



UNIVERSIDAD AUTÓNOMA DE MADRID

Facultad de Ciencias

Departamento de Biología Molecular

## **APLICACIÓN DE LA SECUENCIACIÓN MASIVA DE DNA AL DIAGNÓSTICO DE LOS DEFECTOS CONGÉNITOS DE GLICOSILACIÓN Y DE GLUCOGENOSIS**

Celia Medrano Rodríguez

Tesis Doctoral

Madrid, 2017

Universidad Autónoma de Madrid

Facultad de Ciencias

Departamento de Biología Molecular

## **APLICACIÓN DE LA SECUENCIACIÓN MASIVA DE DNA AL DIAGNÓSTICO DE LOS DEFECTOS CONGÉNITOS DE GLICOSILACIÓN Y DE GLUCOGENOSIS**

Memoria presentada por la Licenciada en Química y Bioquímica

**Celia Medrano Rodríguez**

Para optar al grado de Doctor en Bioquímica, Biología Molecular, Biomedicina y Biotecnología por la  
Universidad Autónoma de Madrid.

Trabajo dirigido por:

**Dra. Celia Pérez-Cerdá**

Dra. en Ciencias Químicas, especialista en Bioquímica Clínica

**Dra. Belén Pérez González**

Profesora Titular del Departamento de Biología Molecular de la Universidad Autónoma de Madrid

Este trabajo ha sido realizado en el Departamento de Biología Molecular, Centro de Biología Molecular Severo Ochoa (CSIC-UAM) en colaboración con el Centro de Diagnóstico de Enfermedades Moleculares (CEDEM) y en el departamento de *Human Genetics* del *University Hospital Gasthuisberg, Katholieke Universiteit Leuven* (Bélgica) bajo la supervisión del Dr. Gert Matthijs. Este trabajo ha sido posible gracias al proyecto del Instituto de Salud Carlos III (PI11/01254 a CPC), la ayuda de posgrado de la Asociación Española para el Estudio de Metabolopatías Congénitas (AEPEMEC) y la ayuda para estancias breves del Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER).



# ÍNDICE

<b>LISTA DE FIGURAS Y TABLAS</b>	<b>1</b>
1 <b>FIGURAS</b>	<b>1</b>
2 <b>TABLAS</b>	<b>2</b>
3 <b>ANEXOS</b>	<b>2</b>
<b>ABREVIATURAS</b>	<b>3</b>
<b>GLOSARIO DE TÉRMINOS ESPECÍFICOS DE NGS</b>	<b>5</b>
<b>RESUMEN / ABSTRACT</b>	<b>7</b>
<b>INTRODUCCIÓN</b>	<b>9</b>
1 <b>ENFERMEDADES    METABÓLICAS    HEREDITARIAS.    DEFECTOS    CONGÉNITOS    DE</b>	
<b>GLICOSILACIÓN Y GLUCOGENOSIS</b>	<b>11</b>
1.1      Defectos Congénitos de Glicosilación	11
1.1.1    Etiología y bases moleculares de la enfermedad	11
1.1.2    Presentación clínica y fisiopatología	14
1.1.3    Diagnóstico	15
1.2      Glucogenosis	17
1.2.1    Etiología y bases moleculares de la enfermedad	17
1.2.2    Presentación clínica y fisiopatología	17
1.2.3    Diagnóstico	19
2 <b>NUEVAS TÉCNICAS DE DIAGNÓSTICO GENÉTICO: SECUENCIACIÓN MASIVA</b>	<b>21</b>
2.1      Metodología	21
2.2      Procesamiento de datos y parámetros de calidad	23
2.3      La secuenciación masiva de DNA aplicada al diagnóstico genético	25
<b>OBJETIVOS</b>	<b>27</b>
<b>MATERIALES Y MÉTODOS</b>	<b>31</b>
1 <b>MATERIALES</b>	<b>33</b>
1.1      Pacientes y cultivos primarios	33
1.2      Reactivos, tampones y soluciones	33
1.3      Anticuerpos	36
1.4      Oligonucleótidos	37
1.5      Equipos	40
1.6      Soporte informático	41

<b>2</b>	<b>MÉTODOS</b>	<b>44</b>
2.1	Análisis bioquímico	44
2.1.1	Cuantificación de las isoformas de transferrina por HPLC	44
2.1.2	Isoelectroenfoque de transferrina, transferencia y revelado	45
2.1.3	Isoelectroenfoque de apolipoproteína CIII, transferencia y revelado	46
2.1.4	Análisis de N-glicanos de proteínas plasmáticas por MALDI-TOF/MS	48
2.2	Estudios celulares	49
2.2.1	Separación electroforética de proteínas en geles SDS-PAGE	49
2.2.2	Electrotransferencia e inmunodetección de proteínas	49
2.2.3	Mantenimiento de los cultivos celulares	50
2.2.4	Detección de proteínas por inmunofluorescencia	50
2.2.5	Evaluación del transporte retrógrado AG-RE	50
2.2.6	Estudios de colocalización	51
2.2.7	Análisis ultraestructural del RE por microscopía electrónica	51
2.3	Análisis genético	52
2.3.1	Aislamiento de ácidos nucleicos y cuantificación	52
2.3.2	Amplificación de ácidos nucleicos	52
2.3.3	Purificación de los productos de PCR y secuenciación por Sanger convencional	53
2.3.4	Secuenciación masiva: Paneles de genes (TES)	53
2.3.5	Secuenciación masiva: Exoma completo (WES)	55
2.3.6	Secuenciación masiva: Exoma clínico (CES)	55
2.4	Soporte informático y análisis <i>in-silico</i>	57
	<b>RESULTADOS</b>	<b>59</b>
<b>1</b>	<b>SELECCIÓN Y ESTRATIFICACIÓN BIOQUÍMICA DE PACIENTES</b>	<b>61</b>
1.1	Selección de pacientes con sospecha de CDG	61
1.2	Selección de pacientes con sospecha de glucogenosis	63
<b>2</b>	<b>EVALUACIÓN DE DISTINTAS APROXIMACIONES DE SECUENCIACIÓN MASIVA</b>	<b>64</b>
2.1	Análisis de variantes detectadas por secuenciación masiva	64
2.1.1	Análisis de variantes en genes candidatos	64
2.1.2	Búsqueda de nuevos genes no asociados a patología	66
2.2	Paneles de captura personalizados (TES)	67
2.2.1	Panel personalizado de genes CDG	67
2.2.2	Panel de captura de genes de glucogenosis	67

2.3	Exoma completo (WES)	68
2.4	Exoma clínico (CES)	69
2.5	Comparativa entre las distintas aproximaciones de NGS empleadas	69
2.6	Detección de grandes deleciones	73
2.7	Análisis de lecturas de genes con pseudogenes	74
<b>3</b>	<b>APLICACIÓN DE LA NGS AL DIAGNÓSTICO DE PACIENTES</b>	<b>76</b>
3.1	Pacientes con defectos congénitos de glicosilación	76
3.1.1	Identificación de un nuevo gen CDG	76
3.1.2	Espectro genético y mutacional de pacientes CDG	78
3.2	Pacientes con glucogenosis. Espectro genético y mutacional	84
3.3	Hallazgos inesperados	88
	<b>DISCUSIÓN</b>	<b>91</b>
1.	<b>EVALUACIÓN DE LAS DISTINTAS APROXIMACIONES DE SECUENCIACIÓN MASIVA</b>	<b>93</b>
2.	<b>APLICACIÓN DE LA NGS AL DIAGNÓSTICO GENÉTICO DE PACIENTES</b>	<b>98</b>
2.1	Espectro genético y mutacional de pacientes CDG	98
2.2	Espectro genético y mutacional de pacientes de glucogenosis	102
2.3	Hallazgos inesperados	103
	<b>CONCLUSIONES</b>	<b>109</b>
	<b>BIBLIOGRAFÍA</b>	<b>113</b>
	<b>ANEXOS</b>	<b>131</b>
	<b>PUBLICACIONES</b>	<b>141</b>
1	<b>PUBLICACIONES QUE HAN DADO LUGAR A ESTA TESIS</b>	<b>143</b>
2	<b>OTROS TRABAJOS PUBLICADOS DURANTE LA ETAPA PREDOCTORAL</b>	<b>200</b>

## 1 FIGURAS

<b>Fig.1</b>	Esquema de la ruta de síntesis de N-glicanos	13
<b>Fig.2</b>	Ruta de síntesis y degradación del glucógeno	18
<b>Fig.3</b>	Etapas de la preparación de la muestra para secuenciación masiva	22
<b>Fig.4</b>	Esquema del procesamiento de datos de NGS	23
<b>Fig.5</b>	Medida del %CDT mediante HPLC	45
<b>Fig.6</b>	Distintos patrones de Trf detectables por IEF	46
<b>Fig.7</b>	Distintos patrones de ApoCIII detectables por IEF	47
<b>Fig.8</b>	Preparación de las muestras para TES	54
<b>Fig.9</b>	Preparación de las muestras para CES	56
<b>Fig.10</b>	Flujo de trabajo empleado para la clasificación de pacientes CDG	62
<b>Fig.11</b>	Protocolo para el análisis de variantes de secuenciación masiva	66
<b>Fig.12</b>	Profundidad de cobertura del exón 5 del gen <i>PMM2</i> del P20	68
<b>Fig.13</b>	Porcentaje de pacientes resueltos por secuenciación masiva	71
<b>Fig.14</b>	Filtrado de variantes obtenidas por secuenciación masiva	72
<b>Fig.15</b>	<i>Allele dropout</i> en la secuenciación por Sanger convencional	73
<b>Fig.16</b>	Región delecionada en el P6	74
<b>Fig.17</b>	Alineamiento del gen <i>ALG1</i> con genes y pseudogenes homólogos a su secuencia	75
<b>Fig.18</b>	Caracterización molecular de CCDC115	77
<b>Fig.19</b>	Espectro genético de los pacientes CDG	78
<b>Fig.20</b>	Mutaciones identificadas en los genes implicados en la N-glicosilación de proteínas	79
<b>Fig.21</b>	Análisis de biomarcadores y estudios moleculares del P11	81
<b>Fig.22</b>	Espectro genético de los pacientes con glucogenosis	84
<b>Fig.23</b>	Mutaciones identificadas en los genes implicados en el metabolismo del glucógeno	85
<b>Fig.24</b>	Interrelación del metabolismo de algunos monosacáridos	107

## 2 TABLAS

<b>Tabla 1</b>	Tipos de glicosilación en los que se han descrito enfermedades	12
<b>Tabla 2</b>	Tampones y soluciones	33
<b>Tabla 3</b>	Reactivos	34
<b>Tabla 4</b>	Anticuerpos	36
<b>Tabla 5</b>	Oligonucleótidos	37
<b>Tabla 6</b>	Equipos	40
<b>Tabla 7</b>	Programas y bases de datos	41
<b>Tabla 8</b>	Etapas del isoelectroenfoque de transferrina	45
<b>Tabla 9</b>	Etapas del isoelectroenfoque de apolipoproteína CIII	47
<b>Tabla 10</b>	Comparativa entre las aproximaciones de secuenciación masiva utilizadas	57
<b>Tabla 11</b>	Datos clínicos y porcentaje de glicanos anómalos de proteínas plasmáticas de pacientes CDG-II	63
<b>Tabla 12</b>	Comparativa de la profundidad de cobertura media de los genes CDG analizados mediante TES-CDG, WES y CES	70
<b>Tabla 13</b>	Comparativa de las coberturas medias de los genes GSD analizados mediante TES-EMH y CES	71
<b>Tabla 14</b>	Análisis de la profundidad de cobertura para la detección de grandes deleciones	74
<b>Tabla 15</b>	Resumen de datos clínicos, bioquímicos y genéticos de los pacientes CDG	82
<b>Tabla 16</b>	Resumen de datos clínicos y genéticos de los pacientes con glucogenosis	86
<b>Tabla 17</b>	Datos clínicos y espectro mutacional de pacientes que portan mutaciones en genes no asociados a CDG o glucogenosis	89

## 3 ANEXOS

<b>Anexo 1</b>	Relación de genes implicados en los defectos en la N-glicosilación y en los defectos combinados	133
<b>Anexo 2</b>	Relación de genes implicados en la glucogenosis.	135
<b>Anexo 3</b>	Genes capturados en el panel de CDG (TES-CDG)	136
<b>Anexo 4</b>	Genes de glucogenosis capturados en el panel metabólico (TES-EMH) y priorizados en el análisis de variantes de CES	137
<b>Anexo 5</b>	Listado priorizado de genes CDG utilizado para el análisis por CES y WES	138



<b>%CDT</b>	Porcentaje de transferrina deficiente en carbohidratos
<b>ACN</b>	Acetonitrilo
<b>AG</b>	Aparato de Golgi
<b>ALT</b>	Alanina transaminasa
<b>ApoCIII</b>	Apolipoproteína CIII
<b>Asn</b>	Asparagina
<b>AST</b>	Aspartato aminotransferasa
<b>BAM</b>	Formato de almacenamiento de grandes alineamientos de secuencias
<b>BFA</b>	Brefeldin A
<b>BSA</b>	Albúmina de suero bovino ( <i>Bovine Serum Albumin</i> )
<b>CDG</b>	Defectos congénitos de glicosilación ( <i>Congenital Disorders of Glycosylation</i> )
<b>CES</b>	Panel de genes asociados a enfermedad anotados en OMIM ( <i>Clinical Exome Sequencing</i> )
<b>CK</b>	Creatina kinasa
<b>CNV</b>	Cambio en el número de copias ( <i>Copy Number Variation</i> )
<b>COG</b>	Complejo oligomérico de Golgi ( <i>Conserved Oligomeric Golgi</i> )
<b>DC</b>	Profundidad de cobertura ( <i>Depth of Coverage</i> )
<b>DMSO</b>	Dimetil sulfóxido
<b>DTT</b>	1,4-ditio-L-treitol
<b>EDTA</b>	Ácido etilen diamino tetraacético
<b>ECL</b>	Sistema de revelado por quimioluminiscencia ( <i>Enhanced Chemiluminescence</i> )
<b>FBS</b>	Suero bovino fetal ( <i>Fetal Bovine Serum</i> )
<b>Fuc</b>	Fucosa
<b>Gal</b>	Galactosa
<b>Glc</b>	Glucosa
<b>GlcA</b>	Ácido glucurónico
<b>Gln</b>	Glucosamina
<b>GM130</b>	Proteína de la matriz del cis-Golgi de 130KDa
<b>GSD</b>	Glucogenosis ( <i>Glycogen Storage Disease</i> )
<b>HGMD</b>	Base de datos de mutaciones humanas ( <i>Human Gene Mutation Database</i> )
<b>HFI</b>	Intolerancia hereditaria a la fructosa ( <i>Hereditary Fructose Intolerance</i> )
<b>HPLC</b>	Cromatografía de líquidos de alta resolución
<b>IEF</b>	Isoelectroenfoque
<b>IF</b>	Inmunofluorescencia
<b>Indel</b>	Inserción - delección

<b>LDL</b>	Lipoproteínas de baja densidad ( <i>Low Density Lipoprotein</i> )
<b>LoF</b>	Mutaciones de pérdida de función ( <i>Loss of Function</i> )
<b>MALDI-TOF/MS</b>	Espectrometría de masas de desorción/ionización por láser acoplado a un detector de tiempo de vuelo
<b>Man</b>	Manosa
<b>MEM</b>	Medio mínimo esencia de Eagle
<b>Mb</b>	Mil pares de bases (mega-pares de bases)
<b>NAcGal</b>	N-acetilgalactosamina
<b>NAcGlc</b>	N-acetilglucosamina
<b>NGS</b>	Secuenciación masiva ( <i>Next-Generation Sequencing</i> )
<b>O/N</b>	Durante toda la noche ( <i>Over Night</i> )
<b>pb</b>	Par de bases
<b>PBS</b>	Tampón fosfato salino ( <i>Phosphate Buffered Saline</i> )
<b>RE</b>	Retículo endoplasmático
<b>rpm</b>	Revoluciones por minuto
<b>RT-PCR</b>	<i>Reverse transcription polymerase chain reaction</i>
<b>SDS</b>	Dodecilsulfato sódico ( <i>Sodium Dodecyl Sulfate</i> )
<b>SDS-PAGE</b>	Electroforesis en geles de poliacrilamida en presencia de SDS
<b>Ser</b>	Serina
<b>SNP</b>	Polimorfismos de un único nucleótido ( <i>Single Nucleotide Polymorphism</i> )
<b>SNV</b>	Variantes de un único nucleótido ( <i>Single Nucleotide Variant</i> )
<b>TBS</b>	Tampón Tris salino ( <i>Tris-Buffered Saline</i> )
<b>TES</b>	Panel de genes personalizado ( <i>Targeted Exome Sequencing</i> )
<b>TFA</b>	Ácido trifluoroacético
<b>Trf</b>	Transferrina
<b>UTR</b>	Región no traducida ( <i>Untranslated Region</i> )
<b>VCF</b>	Formato de almacenamiento de variantes obtenidas por NGS ( <i>Variant Calling File</i> )
<b>WB</b>	Western blot
<b>WES</b>	Secuenciación del exoma completo ( <i>Whole-Exome Sequencing</i> )
<b>WGS</b>	Secuenciación del genoma completo ( <i>Whole-Genome Sequencing</i> )
<b>Xyl</b>	Xilosa

Algunos términos en inglés ampliamente utilizados en Biología Molecular y sin clara traducción en castellano se muestran en cursiva.

<b>Adaptador</b>	DNA sintético que contiene secuencias complementarias a los oligonucleótidos que se emplean en la amplificación y en la secuenciación de la librería y el índice.
<b>Anotación de variantes</b>	Integración de la información de las variantes obtenida de bases de datos contrastadas.
<b>Cobertura horizontal</b>	Número total de pares de bases de las regiones de interés a secuenciar.
<b><i>Clinical exome sequencing (CES)</i></b>	Panel comercial de 4.813 genes asociados a patología en humanos y anotados en la base de datos del OMIM.
<b><i>Clustering</i></b>	Amplificación clonal y agrupamiento de los <i>template</i> de la librería.
<b><i>Demultiplexing</i></b>	Identificación de la muestra de origen de una lectura a partir de su índice.
<b>Índice</b>	6-8pb del adaptador que se utilizan como código de barras identificativo de la muestra.
<b>Lecturas o reads</b>	Secuencias obtenidas del secuenciador.
<b>Librería</b>	Conjunto de <i>templates</i> .
<b><i>Paired-end</i></b>	Secuenciación de un fragmento de DNA a partir de sus dos extremos, de manera que se generan lecturas <i>forward</i> y <i>reverse</i> .
<b>Profundidad de cobertura</b>	Número de veces que cada nucleótido del genoma está representado en las lecturas.
<b><i>Strand-bias</i></b>	Situación que se produce cuando una variante está representada exclusivamente por lecturas de un único tipo ( <i>forward</i> o <i>reverse</i> ).
<b><i>Targets</i></b>	Regiones de interés.
<b><i>Template</i></b>	Fragmento de DNA que tiene unido los adaptadores.
<b><i>Trimming</i></b>	Eliminación de la secuencia del adaptador de la lectura.
<b><i>Variant calling</i></b>	Detección de variantes en las lecturas que difieren respecto de la secuencia de referencia.
<b>VUS</b>	Variante de significado incierto ( <i>variant of unknown significance</i> ).
<b><i>Whole-exome sequencing (WES)</i></b>	Secuenciación del exoma completo.

Up to now, genetic analysis of human inherited diseases is based on gene-by-gene Sanger sequencing. In locus-heterogeneous disorders, genes are sequenced according to their population frequency after performing time-consuming cellular and/or biochemical assays in an attempt to reduce to a single or a small group of candidate genes to be analyzed.

In order to improve the genetic diagnosis of congenital disorders of glycosylation (CDG) and glycogen storage diseases (GSD), two locus-heterogeneous disorders affecting the glycosylation pathways or the enzymes involved in the glycogen metabolism respectively, we applied the cutting-edge DNA next-generation sequencing (NGS) technology. We have evaluated three different NGS-based tests to analyze 39 CDG and 47 GSD suspected patients. First, we captured the exome of targeted genes (TES) previously described to be associated to CDG or GSD. Diagnosis rate of the customized CDG panel was low probably due to the limited number of genes captured of this emerging disorder. In order to improve the diagnosis rate of both diseases, we captured the whole exome (whole-exome sequencing, WES) or the exome of genes associated with human genetic diseases as annotated in OMIM database (clinical exome sequencing, CES). Both tests were used to broaden the number of analyzable genes beyond those already described, allowing the detection of mutations in non-CDG or GSD-associated genes causing symptoms that might overlap with them. Besides, using WES we identified mutations in a new gene not associated to any disease so far.

Additionally, NGS showed higher sensibility than Sanger sequencing since it avoided allele dropout in two samples and allowed the detection of a copy number variation in *ALG1*. Moreover, it is noteworthy the accuracy and sensibility of the technology in the detection of variants located in genes with high sequence-homology, such as *ALG1*.

Overall, we reached a genetic diagnosis of approximately 50% of the patients analyzed. In CDG, we identified 23 variants, 12 of them novel, in 13 previously described CDG-genes. Furthermore, we described a new CDG-gene not associated to any human pathology so far (*CCDC115*). Comprehensive functional studies performed on the patient-derived fibroblasts indicated that the protein codified by *CCDC115* is probably involved in the Golgi homeostasis, impairing protein glycosylation. In GSD we identified 23 changes, 10 of them novel, in 6 already-described genes. Besides, we identified an overall of six patients bearing mutations in genes not associated to the initially suspected disease. The accurate genetic diagnosis allowed in five patients to undergo an accurate tailored treatment improving their outcome.

Summing up, our results highlight the usefulness of NGS to be applied to the diagnosis of two locus-heterogeneous disorders. It is noteworthy that an accurate genetic diagnosis is needed to provide a genetic counseling, to prescribe tailored treatments and to provide insights to guide research towards new therapies based on the mechanism of actions of the mutations identified (precision medicine).

# INTRODUCCIÓN

<b>1</b>	<b>ENFERMEDADES METABÓLICAS HEREDITARIAS. DEFECTOS CONGÉNITOS DE GLICOSILACIÓN Y GLUCOGENOSIS</b>	<b>11</b>
1.1	Defectos Congénitos de Glicosilación	11
1.1.1	Etiología y bases moleculares de la enfermedad	11
1.1.2	Presentación clínica y fisiopatología	14
1.1.3	Diagnóstico	15
1.2	Glucogenosis	17
1.2.1	Etiología y bases moleculares de la enfermedad	17
1.2.2	Presentación clínica y fisiopatología	17
1.2.3	Diagnóstico	19
<b>2</b>	<b>NUEVAS TÉCNICAS DE DIAGNÓSTICO GENÉTICO: SECUENCIACIÓN MASIVA</b>	<b>21</b>
2.1	Metodología	21
2.2	Procesamiento de datos y parámetros de calidad	23
2.3	La secuenciación masiva de DNA aplicada al diagnóstico genético	25

## 1 ENFERMEDADES METABÓLICAS HEREDITARIAS. DEFECTOS CONGÉNITOS DE GLICOSILACIÓN Y GLUCOGENOSIS

Las enfermedades metabólicas hereditarias (EMH) se definen como defectos de origen genético que afectan a la estructura y/o función del producto génico generando una serie de alteraciones en la homeostasis celular y que finalmente producen un impacto sobre el fenotipo (Morava et al., 2015). Su frecuencia es muy baja (<0,05% afectados según EURORDIS, *European Organization of Rare Diseases*), por lo que han sido englobadas dentro del grupo de enfermedades raras.

De entre todas las reacciones que conforman el metabolismo, en este trabajo centraremos la atención en el estudio de los defectos congénitos de glicosilación y los defectos en el metabolismo del glucógeno (glucogenosis).

### 1.1 Defectos Congénitos de Glicosilación

#### 1.1.1 Etiología y bases moleculares de la enfermedad

Los defectos congénitos de glicosilación (CDG, *Congenital Disorders of Glycosylation*) son un grupo heterogéneo de enfermedades genéticas que comprende defectos en las proteínas implicadas en las rutas de glicosilación. Se han descrito defectos genéticos en la síntesis de glicolípidos, en la N-glicosilación de proteínas, en cinco tipos de O-glicosilación de proteínas clasificados en función del primer amino-azúcar que se une (Wopereis et al., 2006) y en múltiples rutas de glicosilación, también denominados defectos combinados (Hennet, 2012) (**Tabla 1**). Durante el presente trabajo nos hemos centrado en los defectos en la N-glicosilación de proteínas y en los defectos combinados (**Figura 1**).

Las deficiencias en las proteínas que participan en etapas comunes a varias rutas de glicosilación se agrupan bajo el nombre de defectos combinados. Se pueden subclasificar en tres grandes grupos: proteínas implicadas en la síntesis de monosacáridos unidos a nucleótidos, proteínas de la biosíntesis de dolicol-fosfato y proteínas de la homeostasis del aparato de Golgi (AG). Los monosacáridos unidos a nucleótidos se sintetizan en el citosol o en el núcleo y actúan como donadores de azúcares en todas las rutas de glicosilación (Freeze and Elbein, 2009). De la misma manera el dolicol-fosfato, un lípido de la membrana del retículo endoplasmático (RE) sobre el que se unen azúcares, actúa como donador de azúcares en el lumen del RE o también como lípido de anclaje donde comienza a ensamblarse el glicano para su posterior traslocación a la proteína o lípido (Buczkowska et al., 2014; Denecke and Kranz, 2009). Cabe resaltar que todas las rutas de glicosilación confluyen en el aparato de Golgi, un orgánulo dinámico organizado en cisternas de las que salen y se fusionan vesículas constantemente. La composición asimétrica en cada cisterna es mantenida gracias a la intervención de proteínas de tráfico vesicular como las proteínas que

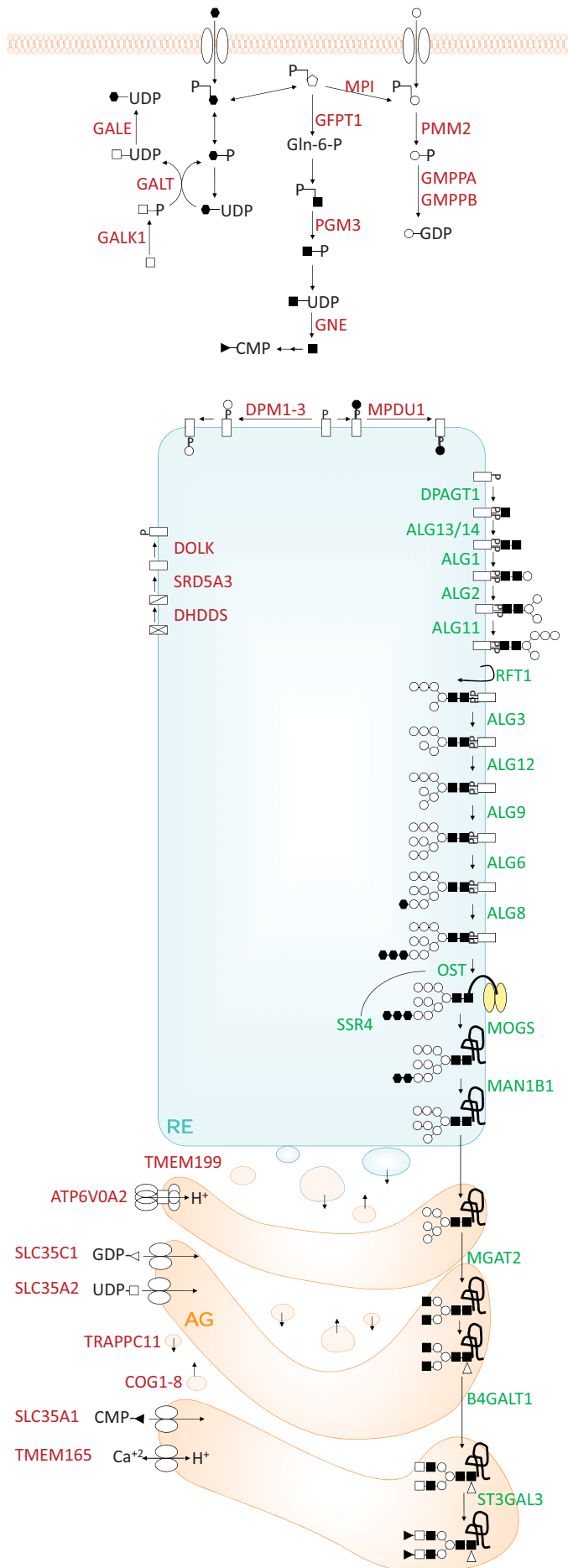
conforman la cubierta de las vesículas COP-I o COP-II o el complejo oligomérico de Golgi (COG) (Smith and Lupashin, 2008; Ungar et al., 2006). En la homeostasis del AG también participan otras proteínas como los transportadores específicos de azúcares unidos a nucleótidos o de iones así como la bomba de protones encargada de mantener el diferente pH de cada cisterna. La deficiencia en cualquiera de estas proteínas afectará al correcto funcionamiento del AG y, con ello, a todas las rutas de glicosilación (Foulquier, 2008; Rivinoja et al., 2012; Rosnoblet et al., 2012).

La N-glicosilación comienza con el ensamblaje del glicano en el RE donde se van uniando azúcares al dolicol-fosfato hasta generar un oligosacárido estándar ( $\text{Glc}_3\text{Man}_9\text{NAcGlc}_2$ ) denominado LLO (*lipid-linked oligosaccharide*). Este oligosacárido estándar es transferido en bloque a la asparragina de la proteína naciente. La unión covalente del glicano ocurre a través del grupo amida de un residuo de asparragina (Asn) ubicado en una secuencia consenso Asn-X-Ser/Thr, siendo X cualquier aminoácido excepto prolina (Park and Zhang, 2011). El glicano unido a la proteína es posteriormente procesado en el RE y el AG donde se eliminan y añaden otros azúcares generando estructuras diversas y especializadas (Jaeken, 2010).

**Tabla 1. Tipos de glicosilación en los que se han descrito enfermedades.**

Glicosilación	Enlace	Síntesis	Funciones
<b>Lípidos</b>			
Glicoesfingolípidos	Ac glucorónico - ceramida	RE y AG	Señalización, reconocimiento célula-célula
Glicofosfatidilinositol	Ác glucorónico-inositol	RE y AG	Hematopoyesis, tráfico de proteínas
<b>Proteínas</b>			
N-glicosilación	NAcGlc - Asn	RE y AG	Plegamiento y estabilidad de proteínas, formación de complejos, señalización y reconocimiento célula-célula
O-manosilglicosilación	Man - Ser/Thr	RE y AG	Estabilidad y función del complejo glicoprotéico distrofina
O-galactosilglicosilación	Gal-hLys	RE y AG	Estructura y estabilidad del colágeno y reconocimiento y activación de sus receptores.
O-xilosilglicosilación	Xyl - Ser	AG	Amortiguamiento mecánico, co-receptor de factores de crecimiento y unión a citoquinas
O-N-acetilgalactosaminil-glicosilación	NAcGal - Ser/Thr	AG	Lubricante, barrera contra patógenos, tráfico de leucocitos y linfocitos
O-fucosilglicosilación	Fuc - Ser/Thr	RE	Hematopoyesis, señalización

Asn, asparragina; Fuc, fucosa; Gal, galactosa; hLys, hidroxilisina; Man, manosa; NAcGal, N-acetilgalactosamina; NAcGlc, N-acetilglucosamina; RE, retículo endoplasmático; Ser, serina; Thr, treonina; Xyl, xilosa. Modificada de (Freeze, 2006).



Se estima que hay entre 250 y 500 genes directamente implicados en la glicosilación (glicogenes) (Schachter and Freeze, 2009). Los defectos en la N-glicosilación y los defectos combinados descritos hasta la fecha se encuentran recogidos en el **anexo 1**.

**Fig 1. Esquema de la ruta de síntesis de N-glicanos.**

Se indican aquellos genes cuyo defecto se ha asociado a enfermedad; en verde N-glicosilación y en rojo los defectos combinados. Los azúcares unidos a nucleótidos se sintetizan en el citosol o núcleo y son utilizados como donadores de azúcares durante toda la ruta. El glicano comienza a ensamblarse sobre un lípido de la membrana del RE hasta la formación de un glicano estándar que es transferido en bloque a la proteína naciente y que posteriormente es modificado en el RE y el AG.

◇ fructosa, ○ glucosa, □ galactosa, Gln glucosamina, ■ N-acetilglucosamina, △ fucosa, ▲ ácido siálico, o manosa, P fosfato, S azufre, □ dolicol, ▭ poliprenol.



### 1.1.2 Presentación clínica y fisiopatología

Los defectos en la N-glicosilación de proteínas (N-CDG) se clasifican en dos grupos, CDG tipo I (CDG-I) que incluye defectos en la síntesis o transferencia del oligosacárido estándar a la proteína, y CDG tipo II (CDG-II) que engloba los defectos en el procesamiento del glicano una vez unido a la proteína. Se nombran con el nombre del gen mutado seguido de la abreviatura CDG (Jaeken et al., 2009). La mayoría siguen un modo de herencia autosómico recesivo, aunque también se han identificado ligadas al X (Bissar-Tadmouri et al., 2014; Ng et al., 2013).

Según FEDER, la frecuencia de estas enfermedades oscila entre 1/50.000 – 1/100.000. La causa más frecuente de CDG es la debida a la deficiencia en *PMM2* (PMM2-CDG, MIM 212065), con más de 700 pacientes descritos en el mundo (Jaeken, 2010) y una incidencia de 1/50.000 (Hennet, 2012). Sin embargo, en otras deficiencias, como en MOGS-CDG, sólo se ha descrito un único caso en el mundo (Praeter et al., 2000).

La N-glicosilación afecta al plegamiento y estabilidad de la proteína, a la formación de complejos proteicos, a la resistencia frente a la acción de proteasas, al transporte inter- e intra-celular, al reconocimiento célula-célula y a la señalización. Se cree que el 50% de las proteínas humanas están N-glicosiladas y están implicadas en una gran cantidad de funciones como el sistema inmune, la determinación del grupo sanguíneo, la respuesta inflamatoria, la coagulación y la protección frente a estrés oxidativo, entre otras (Richard et al., 2009; Wolfe and Krasnewich, 2013).

El impacto sobre las proteínas de una glicosilación aberrante y la gran cantidad de funciones en las que están implicadas justifica el amplio espectro fenotípico de los pacientes CDG. Por lo general, los pacientes CDG debido a deficiencias en la ruta de la N-glicosilación presentan un fenotipo clínico multisistémico con afectación cerebral, muscular, hepática, cardíaca y ocular así como problemas de coagulación (Jaeken, 2010; Scott et al., 2014). Si bien, algunos presentan características fenotípicas particulares (Lefeber et al., 2011; Scott et al., 2014).

### 1.1.3 Diagnóstico

El amplio espectro fenotípico de los pacientes N-CDG hace que sean difícilmente reconocibles, por lo que para su identificación inicial es necesario realizar un cribado a todos los pacientes con sospecha de enfermedad genética que presenten una clínica multisistémica con afectación neurológica de origen inexplicable.

En la actualidad, el cribado de las N-CDG se realiza mediante el análisis de las isoformas sialiladas de la N-glicoproteína plasmática transferrina (Trf) por isoelectroenfoque (IEF) (Hackler et al., 1995), electroforesis capilar (Kuroda et al., 2013) o HPLC (Helander et al., 2010). Estos métodos se basan en la separación de las isoformas de la transferrina plasmática en base a su carga, que es determinada por el número de residuos de ácido siálico localizados en el extremo terminal del glicano N-unido. En general, se pueden distinguir dos patrones anómalos:

- Tipo 1, caracterizado por un incremento en las formas disialo- y asialotransferrina, que apunta a un defecto en la síntesis o transferencia del oligosacárido estándar a la proteína (CDG-I).
- Tipo 2, con un incremento variable en las formas tri-, di-, mono y asialiladas de la transferrina y que apunta a un defecto en el procesamiento del glicano unido a la proteína (CDG-II).

No obstante, se han descrito alteraciones en el perfil de isoformas de Trf debido a otras patologías hereditarias como la galactosemia (Charlwood et al., 1998) y la intolerancia a la fructosa (Jaeken et al., 1996), que pueden considerarse CDG secundarios. Asimismo, se han descrito otras situaciones que podrían originar un perfil alterado tales como patologías hepáticas (Bean et al., 1995; Murawaki et al., 1997), el síndrome hemolítico urémico por infección con *Streptococcus* (de Loos et al., 2002), la presencia de bacterias con actividad sialidasa y en menores de tres meses por inmadurez de los glicanos unidos a las proteínas (Thiel et al., 2012). También cabe destacar que se han descrito perfiles normales en algunos pacientes CDG (Lefeber et al., 2011; Martinez-Duncker et al., 2005; Mohamed et al., 2011; Morava et al., 2009; Stray-Pedersen et al., 2014). Además, la interpretación de los resultados puede complicarse por la presencia de polimorfismos en la cadena aminoacídica de la transferrina (Guillard et al., 2011b; Zühlsdorf et al., 2015), lo que hace necesaria la búsqueda de nuevos biomarcadores para el diagnóstico de los CDG (Van Scherpenzeel et al., 2016).

El procedimiento estándar para las muestras con un perfil de isoformas de Trf tipo 2 continúa con el análisis de las isoformas sialiladas de la O-glicoproteína tipo mucina apolipoproteína-CIII (ApoCIII) mediante IEF o HPLC para, de esta manera, discriminar entre un defecto aislado en la N-glicosilación o un defecto combinado en la N- y O-glicosilación de proteínas (Wopereis et al., 2003).

Se han desarrollado otros métodos para evaluar otras etapas de la ruta de N-glicosilación, como el análisis de dolicoles en plasma (Haeuptle et al., 2010) y el análisis de la estructura de los glicanos unidos a dolicol (Hayes and Varki, 1993; Körner et al., 1998; Panneerselvam and Freeze, 1995), a transferrina (Wada, 2016) o a proteínas plasmáticas (Guillard et al., 2009). También es posible medir la actividad enzimática de algunas de las enzimas en fibroblastos o linfocitos del paciente (Senderek et al., 2011; Van Schaftingen and Jaeken, 1995; Würde et al., 2012; Yang et al., 2013).

Debido a lo complicados que son estos métodos, por lo general se tiende a realizar el diagnóstico genético a partir de una sospecha clínica y un perfil alterado de las isoformas de transferrina y/o de ApoCIII. La identificación del gen deficiente en cada caso es fundamental para poder ofrecer un asesoramiento genético y una terapia, si esta existe. En la actualidad, algunas CDG tienen tratamiento tales como MPI-CDG (MIM 602579) (Niehues et al., 1998), SLC35C1-CDG (MIM 266265), PGM1-CDG (MIM 614921) (Morava, 2014; Tegtmeyer et al., 2014), SLC35A2-CDG (MIM 300896) (Dörre et al., 2015) y SLC39A8-CDG (MIM 616721) (Park et al., 2015). Además, actualmente se están investigando otras terapias específicas de mutación como el uso de chaperonas farmacológicas para las mutaciones de plegamiento en el gen *PMM2* (Yuste-Checa et al., 2016).

## 1.2 Glucogenosis

### 1.2.1 Etiología y bases moleculares de la enfermedad

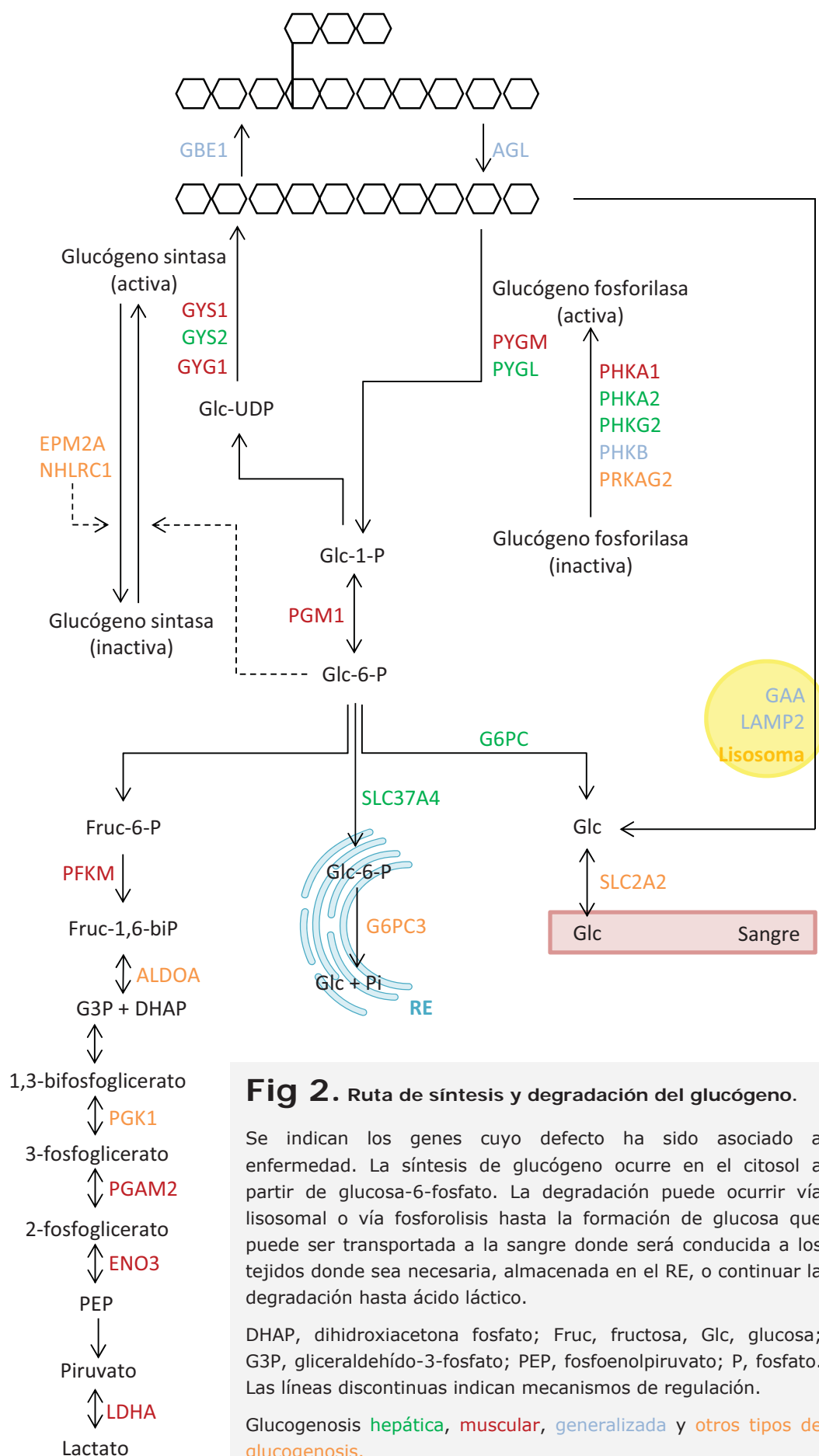
La glucogenosis (GSD, *Glycogen Storage Disease*) es un grupo de enfermedades genéticas hereditarias debida a la deficiencia en alguna de las enzimas implicadas en la síntesis o degradación del glucógeno.

Las células animales almacenan glucosa en forma de glucógeno, principalmente en las células hepáticas y musculares. La síntesis de glucógeno ocurre en el citosol, mientras que la degradación puede ocurrir vía fosforólisis en el citosol o vía lisosomal (**Figura 2**). Ambos procesos están finamente regulados a nivel hormonal ante una situación metabólica determinada del organismo.

Se ha descrito la deficiencia en 27 genes diferentes implicados en el metabolismo del glucógeno que se nombran según el orden cronológico en el que fueron descritos, precedido de la abreviatura GSD. Sin embargo, en la actualidad se tienden a designar según el déficit enzimático. Todas ellas siguen un modo de herencia autosómico recesivo excepto las deficiencias en *LAMP2*, *PGK1*, *PHKA1* y *PHKA2* que están ligadas al X y *PRKAG2* que es autosómica dominante.

### 1.2.2 Presentación clínica y fisiopatología

En la glucogenosis los tejidos más afectados son aquellos en los que el metabolismo del glucógeno juega un papel más importante, el hígado y el músculo. Así pues, las manifestaciones clínicas dependerán de la dificultad que exista en estos tejidos para movilizar sus depósitos de glucógeno. Atendiendo a la expresión clínica y a los hallazgos histopatológicos, se pueden distinguir tres grandes grupos de glucogenosis: hepáticas, musculares y generalizadas, aunque existe un pequeño grupo con una presentación diferente (**Anexo 2**). Si el hígado es el órgano afectado, presentarán hepatomegalia e hipoglucemia además de un retraso estatura-ponderal. En cambio, cuando es el músculo, debutan con debilidad muscular y fatigabilidad precoz al ejercicio e incluso dolor muscular, contracturas y rabdomiolisis cuando el ejercicio es rápido e intenso. Las glucogenosis generalizadas debutan con afectación hepática, muscular y cardíaca, entre las que se encuentran las enzimas de la degradación no fosforolítica del glucógeno GAA y LAMP2, y enzimas ubicuas como PHKB, GBE1 y AGL.



**Fig 2. Ruta de síntesis y degradación del glucógeno.**

Se indican los genes cuyo defecto ha sido asociado a enfermedad. La síntesis de glucógeno ocurre en el citosol a partir de glucosa-6-fosfato. La degradación puede ocurrir vía lisosomal o vía fosforólisis hasta la formación de glucosa que puede ser transportada a la sangre donde será conducida a los tejidos donde sea necesaria, almacenada en el RE, o continuar la degradación hasta ácido láctico.

DHAP, dihidroxiacetona fosfato; Fruc, fructosa; Glc, glucosa; G3P, gliceraldehído-3-fosfato; PEP, fosfoenolpiruvato; P, fosfato. Las líneas discontinuas indican mecanismos de regulación.

Glucogenosis **hepática**, **muscular**, **generalizada** y **otros tipos de glucogenosis**.

La deficiencia en algunas de las enzimas de la glucogenosis como PFKM (MIM 610681), EPM2A (MIM 607566) y NHLRC1 (MIM 608072) y la forma adulta de la deficiencia en GBE1 (MIM 607839) debutan además con afectación neurológica, a diferencia del resto de las glucogenosis. Esto es debido a que estas deficiencias generan un glucógeno anómalo con pocas ramificaciones llamado poliglucosán que agrega y precipita junto con otras proteínas (cuerpos de poliglucosán) en los hepatocitos, miocitos, músculo esquelético y neuronas principalmente (Hays et al., 1981; Lohi et al., 2005; Wierzbza-Bobrowicz et al., 2008).

Además, algunas de las enzimas implicadas en los procesos de síntesis o degradación del glucógeno tienen una presentación tejido-específica diferente. Este es el caso de *ALDOA* (MIM 103850) que se expresa en músculo y eritrocitos y cuya deficiencia cursa con afectación muscular y anemia hemolítica (Kreuder et al., 1996) y la deficiencia en *PRKAG2* (MIM 602743) que debuta con cardiomiopatía hipertrófica al expresarse en miocardiocitos (Blair et al., 2001).

### 1.2.3 Diagnóstico

La deficiencia en la síntesis o degradación del glucógeno se puede sospechar por la presencia de hipoglucemias en ayunos cortos con aumento de ácido láctico y cuerpos cetónicos. En el examen bioquímico también se valora la elevación de enzimas musculares como la creatín kinasa (CK) o enzimas hepáticas como AST y ALT orientando el diagnóstico hacia una glucogenosis muscular, hepática o generalizada. Además, la hiperuricemia es un indicador de nefropatía, característico de las deficiencias en *G6PC* (MIM 613742) y *SLC37A4* (MIM 602671).

También se realizan test dinámicos donde se administra al paciente un compuesto implicado en el metabolismo del glucógeno (glucosa, galactosa, glicerol o fructosa) y se analiza cómo es metabolizado hacia la síntesis de glucógeno y la degradación a láctico. Además, se puede valorar la respuesta del organismo ante una situación determinada, como la respuesta a glucagón en estado postprandial o el test de isquemia en el ejercicio donde se analiza la degradación de glucógeno a glucosa y finalmente a láctico (McArdle and Verel, 1956; Munsat, 1970).

Además, puede tomarse una biopsia hepática o muscular y realizar una evaluación histológica y/o histoquímica del tejido para determinar el aspecto del mismo e identificar acúmulos de glucógeno. En algunos casos la actividad enzimática puede medirse en sangre, biopsia muscular o hepática. No obstante, la confirmación diagnóstica requiere siempre del análisis mutacional del gen y, puesto que para la medida enzimática puede ser necesaria una muestra invasiva, actualmente se tiende a realizar el diagnóstico genético a partir de una sospecha clínica y los resultados de la analítica en sangre (Davit-Spraul et al., 2011).

La identificación del gen deficiente en cada caso resulta crucial para poder indicar unas pautas nutricionales y de ejercicio físico adecuadas (Chen et al., 1993; Heller et al., 2008). Además, los pacientes deficientes en *GAA* (enfermedad de Pompe, MIM 232300) pueden beneficiarse de la terapia de reemplazamiento enzimático comercializada como Myozyme® (Genzyme) (Kishnani et al., 2006).

Un retraso en el diagnóstico de aquellas enfermedades con tratamiento podría resultar en una grave descompensación metabólica con un gran impacto en el fenotipo que puede llevar incluso a la muerte. Esto pone de relieve la importancia de un diagnóstico rápido y preciso.

Tanto los CDG como la glucogenosis debutan con una clínica inespecífica y carecen de un biomarcador patognomónico para su identificación, por lo que es necesario recurrir al diagnóstico genético. Si bien, el diagnóstico genético de estas enfermedades resulta complejo al ser enfermedades genéticamente heterogéneas con un gran número de genes implicados, lo que hace necesaria la implementación de nuevas técnicas de análisis genético.

## 2 NUEVAS TÉCNICAS DE DIAGNÓSTICO GENÉTICO: SECUENCIACIÓN MASIVA

---

El descubrimiento de los dideoxinucleótidos para la secuenciación por terminación de cadena por Sanger marcó un hito en la historia de la secuenciación de DNA (Sanger et al., 1977). Basándose en esta tecnología se han ido desarrollando paulatinamente nuevos métodos de secuenciación automática (Marsh et al., 1997; Smith et al., 1986).

Actualmente, el escenario de la secuenciación de DNA ha cambiado rápidamente tras el desarrollo de la secuenciación masiva (NGS, *Next-Generation Sequencing*). En esencia, la tecnología de NGS permite el análisis de muchas reacciones que se secuencian simultáneamente en una única carrera, por lo que son capaces de proporcionar respuestas a escala genómica en un tiempo corto. El cuello de botella de esta tecnología radica en el análisis de los datos debido a la gran cantidad de información obtenida. El desarrollo de algoritmos de análisis de datos y la interpretación de los resultados en la actualidad son el mayor desafío de la NGS.

### 2.1 Metodología

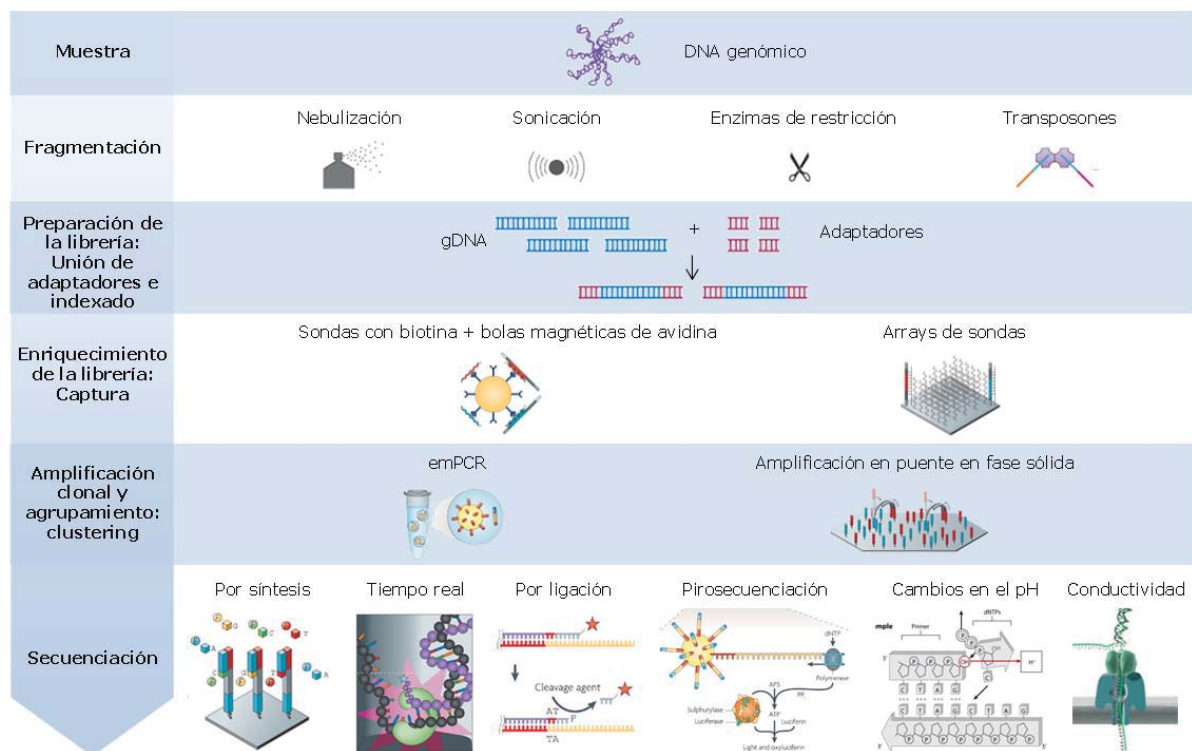
La metodología en la que se basan la mayoría de las aproximaciones de NGS de DNA consiste en una fragmentación de la muestra, preparación de las librerías, amplificación y secuenciación (**Figura 3**). Además, puede realizarse una etapa opcional de captura antes de la amplificación para enriquecer la librería en las regiones de interés. No obstante, la metodología de secuenciación subyacente y de preparación de la librería son propias de cada plataforma comercial. A continuación se detalla cada una de las etapas.

En primer lugar, se parte de una muestra de DNA que se somete a fragmentación al azar y a continuación se une DNA sintético a ambos extremos de cada fragmento. Este DNA sintético denominado “adaptador” contiene secuencias universales complementarias a los oligonucleótidos que se van a emplear posteriormente en la amplificación y en la secuenciación y 6-8pb que se utilizan como código de barras o índice identificativo de la muestra (indexado). Los fragmentos así preparados se llaman *template* y su conjunto forma una librería que, además, puede ser enriquecida en aquellos *template* que contienen las regiones de interés.



El enriquecimiento de la librería o etapa de captura de las regiones de interés se puede realizar con sondas unidas a biotina que se purifican con avidina o con sondas inmovilizadas en un array. Posteriormente, los *template* se amplifican clonalmente y se agrupan (*clustering*) por PCR en emulsión (emPCR) o por amplificación en puente en fase sólida y finalmente se secuencian.

Según la plataforma utilizada, la secuenciación puede ser por síntesis utilizando nucleótidos en cuya base nitrogenada tienen unido un fluoróforo que se detectan durante la copia del *template*, por secuenciación a tiempo real con nucleótidos marcados con fluoróforos en la cadena de fosfatos, por ligación con sondas complementarias marcadas con fluoróforos, por pirosecuenciación donde se registra de manera indirecta el pirofosfato liberado tras la unión del nucleótido complementario, por cambios en el pH producidos al liberarse un protón tras la unión del nucleótido complementario o por cambios en la conductividad que se producen al pasar una hebra de DNA por un nanoporo de una membrana polimérica (Ambardar et al., 2016; Metzker, 2009). No obstante, existen otros métodos de secuenciación masiva alternativos que utilizan otras aproximaciones (Bahassi et al., 2014; Mardis, 2013; Pareek et al., 2011).

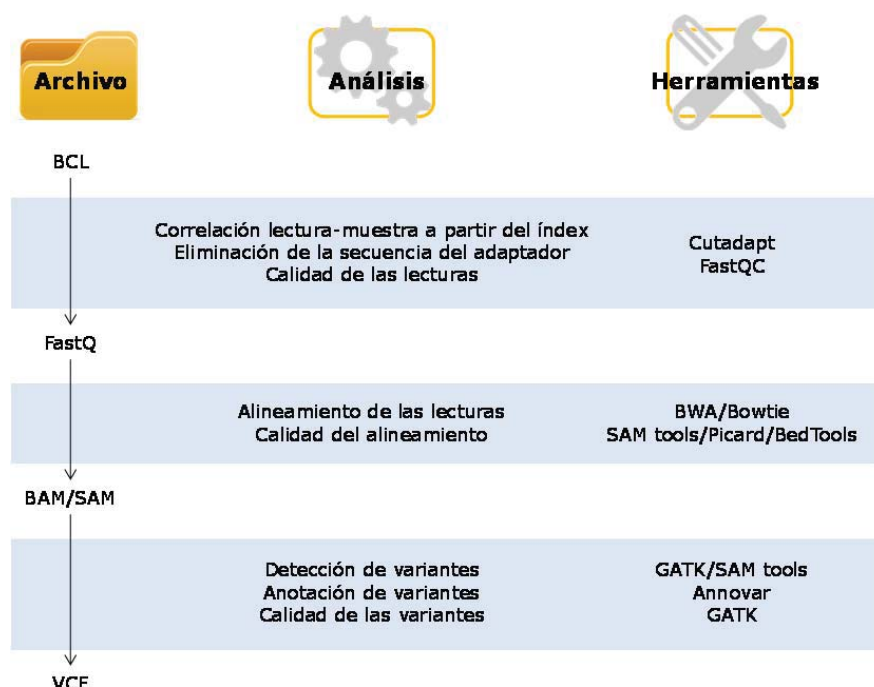


**Fig 3.** Etapas de la preparación de la muestra para secuenciación masiva. A la izquierda sobre una flecha se resume el flujo de trabajo. Las muestras de gDNA se fragmentan y a cada extremo de los fragmentos se unen unos adaptadores (preparación de la librería). La librería se enriquece en las regiones de interés y se amplifica para finalmente ser secuenciada. Adaptada de (Loman et al., 2012; Mardis, 2013; Metzker, 2009).

## 2.2 Procesamiento de datos y parámetros de calidad

La información obtenida del secuenciador contiene el conjunto de secuencias de las muestras analizadas. Cada una de estas secuencias se denomina lectura o *read*. Esta información contenida en un archivo debe ser procesada mediante una serie de herramientas bioinformáticas con el fin de obtener un fichero de variantes para su análisis e interpretación (**Figura 4**).

Las lecturas generadas por el secuenciador en formato BCL son procesadas con el fin de determinar la muestra de origen de cada una de ellas a partir del índice (*demultiplexing*) y eliminar la secuencia correspondiente al adaptador (*trimming*). El archivo obtenido tras este proceso tiene el formato FastQ. Las lecturas son alineadas frente una secuencia de referencia y se exportan a un fichero de datos tipo BAM / SAM (Li et al., 2009). El análisis del archivo BAM permite comparar la secuencia de referencia con la de las lecturas y detectar aquellas variantes que difieren respecto de la referencia (*variant calling*). Estas variantes suelen denotarse como SNV de las siglas en inglés *Single Nucleotide Variant*. La compilación de los datos con algunas bases de datos contrastadas permite obtener información adicional sobre las variantes detectadas (anotación de variantes). El fichero de salida tras este análisis es un VCF y es el que se emplea principalmente en el posterior análisis cuando se integran estos resultados con los datos clínicos y bioquímicos del paciente.



**Fig 4.** Esquema del procesamiento de datos de NGS. La columna de la izquierda indica el tipo de formato del archivo de entrada y salida tras el procesamiento bioinformático realizado en cada caso. La columna central detalla el análisis realizado en cada etapa y a la izquierda la herramienta empleada.

Durante el procesamiento de los datos obtenidos de secuenciación masiva se miden parámetros de calidad a tres niveles distintos:

- Calidad de las bases de las lecturas: Se determina el nivel de fiabilidad de que la base leída sea correcta según la relación señal/ruido. Por lo general, se considera que una base tiene calidad suficiente cuando tiene un 99% de precisión (Ewing et al., 1998).
- Calidad del alineamiento de las lecturas frente al genoma de referencia: Por lo general, la secuenciación se realiza a partir de ambos extremos del *template* de tal manera que para un mismo fragmento tendremos una lectura *forward* y una *reverse* dando lugar a lecturas apareadas o *paired-end*. En la evaluación de la calidad del alineamiento se concatenan las lecturas provenientes de un mismo fragmento y se comprueba que poseen la distancia entre ambas y la orientación esperada. Además, se determina el número de veces que alinea una lectura en el genoma de referencia y si esta está duplicada.
- Calidad de las variantes o SNV: Se estipula que una variante es real cuando está representada en más de un 15% de las lecturas totales. Además, se comprueba que la variante está igualmente representada en lecturas *forward* y *reverse*. En caso contrario, se considera un *strand bias* y se descarta la variante (Park et al., 2014).

Tras el procesamiento bioinformático de los datos obtenidos del secuenciador y haber aplicado los filtros de calidad descritos, se procede al análisis e interpretación de los resultados integrando la información del archivo VCF con los datos clínicos y bioquímicos del paciente. Además, en ocasiones es necesario recurrir al archivo BAM para revisar los alineamientos de las secuencias. El número de veces que está leída una posición del genoma se denomina “profundidad de cobertura” (DC, *depth of coverage*). Por lo general, se consideran una DC mínima de 10-20x para poder detectar SNV en heterocigosis con precisión (Rehm et al., 2013). Ante una variante respecto del genoma de referencia se calcula la relación entre el número de veces que se ha leído la variante frente al DC de esa posición y de esta manera se define si la variante está en homocigosis (>80%) o en heterocigosis (15-80%). Es importante no confundir el término “profundidad de cobertura” con “cobertura horizontal”. Esta última hace referencia al número total de pares de bases de las regiones de interés a secuenciar (*targets*).

Durante el análisis de las variantes se realizan estudios *in-silico* y se consultan bases de datos contrastadas con el fin de poder clasificar las variantes como mutaciones causantes de enfermedad, polimorfismos o variantes de significado incierto (VUS, *variant of unknown significance*). Además, en ocasiones será necesario complementar el análisis con estudios funcionales.

## 2.3 La secuenciación masiva de DNA aplicada al diagnóstico genético

Existen distintas aproximaciones de la secuenciación masiva de DNA según las regiones de interés a secuenciar, los paneles de genes, el exoma o el genoma completo. Cada una de las aproximaciones tiene una aplicabilidad diferente, haciendo que sea una técnica muy versátil (Xue et al., 2015).

En los paneles de genes personalizados (TES, *targeted exome sequencing*) la librería se enriquece en determinadas regiones genómicas a criterio del profesional, como por ejemplo, un conjunto de genes implicados en una misma patología. Están especialmente indicados para el diagnóstico de enfermedades clínica y bioquímicamente bien definidas donde los genes implicados son bien conocidos (Trujillano et al., 2014).

También existen paneles comerciales que permiten el análisis de los exones y 50pb de las regiones intrónicas flanqueantes de los genes asociados a patología en humanos y anotados en la base de datos del OMIM (CES, *clinical exome sequencing*). Estos paneles amplios son muy útiles para el diagnóstico de enfermedades donde varios genes pueden producir un fenotipo similar o para encontrar la causa molecular cuando presenta un fenotipo inespecífico (Okazaki et al., 2016).

Otras aproximaciones más amplias son la secuenciación del exoma completo (WES, *whole-exome sequencing*), que captura los exones y 25-50pb de la región intrónica flanqueante de todos los genes que codifican para proteínas (Rabbani et al., 2014), o incluso la secuenciación del genoma completo (WGS, *whole-genome sequencing*) (van El et al., 2013).

La aplicación de la secuenciación masiva ha supuesto una revolución en el diagnóstico genético ya que permite detectar con precisión prácticamente todo el espectro de variantes genéticas que puedan tener una consecuencia molecular y un impacto sobre el fenotipo (Koboldt et al., 2010). NGS resulta especialmente prometedora para el diagnóstico de enfermedades genéticamente heterogéneas ya que permite obtener un diagnóstico genético rápido al poder analizar muchos genes de manera simultánea. No obstante, el diseño del estudio y el análisis de datos suponen un gran desafío. Además, la implementación de la técnica como método diagnóstico debe superar un proceso de validación donde además, debe seleccionarse la aproximación idónea en cada caso según el objetivo del estudio.

## OBJETIVOS

El diagnóstico genético de los pacientes con enfermedades metabólicas hereditarias es necesario para poder ofrecer un correcto asesoramiento genético que permita facilitar un diagnóstico prenatal y/o preimplantacional, así como proporcionar posibles tratamientos personalizados.

En el caso de las enfermedades genéticamente heterogéneas, como son los defectos congénitos de glicosilación y la glucogenosis, la identificación de los genes y mutaciones causantes de enfermedad mediante la tecnología *gold-standard* utilizada hoy en día, la secuenciación por Sanger convencional, resulta ineficaz, quedando muchos pacientes sin un diagnóstico genético. Esta situación pone de relieve la necesidad de la implementación de nuevas técnicas de análisis genético.

Por ello, el **objetivo principal** de este trabajo ha sido la evaluación de la tecnología de la secuenciación masiva de DNA aplicada al diagnóstico de los defectos congénitos de glicosilación y de la glucogenosis.

Para ello, se establecieron los siguientes **objetivos específicos**:

1. Selección y estratificación bioquímica de pacientes con sospecha de defectos congénitos de glicosilación y de glucogenosis.
2. Evaluación de distintas aproximaciones de secuenciación masiva aplicables al diagnóstico genético: captura de genes específicos, captura del exoma clínico y exoma completo.
3. Evaluación de parámetros de filtrado y desarrollo de un flujo de trabajo para el análisis de variantes detectadas por secuenciación masiva.
4. Descripción del espectro genético y mutacional de pacientes con CDG o glucogenosis.

# MATERIALES Y MÉTODOS

<b>1</b>	<b>MATERIALES</b>	<b>33</b>
1.1	Pacientes y cultivos primarios	33
1.2	Reactivos, tampones y soluciones	33
1.3	Anticuerpos	36
1.4	Oligonucleótidos	37
1.5	Equipos	40
1.6	Soporte informático	41
<b>2</b>	<b>MÉTODOS</b>	<b>44</b>
2.1	Análisis bioquímico	44
2.1.1	Cuantificación de las isoformas de transferrina por HPLC	44
2.1.2	Isoelectroenfoque de transferrina, transferencia y revelado	45
2.1.3	Isoelectroenfoque de apolipoproteína CIII, transferencia y revelado	46
2.1.4	Análisis de N-glicanos de proteínas plasmáticas por MALDI-TOF/MS	48
2.2	Estudios celulares	49
2.2.1	Separación electroforética de proteínas en geles SDS-PAGE	49
2.2.2	Electrotransferencia e inmunodetección de proteínas	49
2.2.3	Mantenimiento de los cultivos celulares	50
2.2.4	Detección de proteínas por inmunofluorescencia	50
2.2.5	Evaluación del transporte retrógrado AG-RE	50
2.2.6	Estudios de colocalización	51
2.2.7	Análisis ultraestructural del RE por microscopía electrónica	51
2.3	Análisis genético	52
2.3.1	Aislamiento de ácidos nucleicos y cuantificación	52
2.3.2	Amplificación de ácidos nucleicos	52
2.3.3	Purificación de los productos de PCR y secuenciación por Sanger convencional	53
2.3.4	Secuenciación masiva: Paneles de genes (TES)	53
2.3.5	Secuenciación masiva: Exoma completo (WES)	55
2.3.6	Secuenciación masiva: Exoma clínico (CES)	55
2.4	Soporte informático y análisis <i>in-silico</i>	57

## 1 MATERIALES

### 1.1 Pacientes y cultivos primarios

En este trabajo se han incluido muestras biológicas de 86 pacientes, 39 con sospecha de CDG y 47 con sospecha de glucogenosis. Para los análisis bioquímicos se ha partido de plasma o suero que se conservó a -20°C hasta su utilización. Para los análisis genéticos se partió de DNA o RNA extraídos de fibroblastos de piel o sangre. Para los estudios celulares se ha utilizado fibroblastos cultivados obtenidos de biopsia de piel. La línea de fibroblastos control empleada, CC02509, fue obtenida de la casa comercial Lonza.

Todo el material biológico de pacientes fue obtenido con el consentimiento informado autorizado por los padres o tutores. Toda la investigación llevada a cabo con este material está autorizada por el Comité de Ética de la Investigación de la Universidad Autónoma de Madrid, respetando los principios fundamentales de la declaración de Helsinki, del Convenio del Consejo de Europa relativo a los derechos humanos y la biomedicina y de la declaración Universal de la UNESCO sobre el Genoma Humano y los derechos humanos.

### 1.2 Reactivos, tampones y soluciones

Tabla 2. Tampones y soluciones.

Solución	Composición
PBS	NaCl 137mM, KCl 2,7mM, Na <sub>2</sub> HPO <sub>4</sub> 8mM, KH <sub>2</sub> PO <sub>4</sub> 1,5mM
Fijador microscopía electrónica	4% paraformaldehído, 2% glutaraldehído en tampón fosfato sódico 0,1M pH 7,4
Solución de bloqueo IF	0,1% tritón X100, 1% BSA, 20% FBS en PBS
Solución de bloqueo WB	PBS o TBS 1x, leche en polvo 5%
Solución de deshibridación	Glicina 0,2M en TBS con Tween 20 0,05%
Solución de lavado WB	TBS 1x, Tween-20 0,05%
Tampón de lisis SDS-PAGE	Tris-HCl 10mM pH 7,5, NaCl 150mM, tritón 0,1%, glicerol 10%
Solución de permeabilización IF	Glicina 0,05mM en PBS
TBS	Tris-HCl 50mM, NaCl 276mM, KCl 54mM pH 8



**Tabla 3. Reactivos.**

Reactivo / Producto	Referencia	Casa comercial	Aplicación
Acetonitrilo (ACN)	100003	Merck	Glicanos MALDI-TOF/MS
Agarose D1 Low EEO	8010	Conda	Electroforesis de productos de PCR
Anfolinas 3,5-5,0	80112589	Amersham	Isoelectroenfoque
Anfolinas 4,2-4,9	17056201	Amersham	Isoelectroenfoque
Beta-mercaptoetanol ( $\beta$ ME)	444203	Merck	Glicanos MALDI-TOF/MS
BigDye® Terminator v3.1 Cycle Sequencing kit	4337455	Applied Biosystems	Secuenciación
BioRad Protein assay (Bradford)	500-0006	BioRad	Western Blot
Brefeldin A (BFA)	B7651	Sigma	Transporte retrógrado AG-RE
Bovine serum albumin (BSA)	A6003	Sigma	Inmunofluorescencia
Citrato de férrico	103863	Merck	Isoelectroenfoque
Cloroformo	102445	Merck	Glicanos MALDI-TOF/MS
Columna set-pack C18	WAT043395	Waters	Glicanos MALDI-TOF/MS
Columnas de grafito	60106-304	Thermo Scientific	Glicanos MALDI-TOF/MS
Cubreobjetos Ø 10mm	0384	Thermo Scientific	Inmunofluorescencia
Ditiotreitol (DTT)	D0632	Sigma	Western Blot
E.Z.N.A.® Cycle Pure Kit	D6492-02	Omega Biotek	Purificación de productos de PCR
ECL Western Blotting Detection	RPN2209	GE Healthcare Life Sciences	Inmunodetección
EDTA-free protease inhibitor cocktail	11836170001	Roche	Western Blot
ExoSAP-IT® PCR Product Cleanup	78201	Affymetrix-Isogen	Purificación de productos de PCR
Fast Start TAQ DNA Polymerase	12032945001	Roche	Amplificación de ácidos nucleicos
Formalina 4%	H-T501128	Sigma	Inmunofluorescencia
Haloplex Custom kit	Personalizado	Agilent	TES
Kit Ready-Prep %CDT by HPLC™	195-6671	BioRad	%CDT
MagNA Pure Compact DNA Isolation Kit-large volume	3730972001	Roche	Aislamiento de ácidos nucleicos
MagNA Pure Compact Nucleic Acid Isolation Kit I	3730964001	Roche	Aislamiento de ácidos nucleicos
MagNA Pure Compact RNA Isolation Kit	4802993001	Roche	Aislamiento de ácidos nucleicos
MagNA Pure LC DNA Isolation Kit-lysis/binding buffer refill	3246752001	Roche	Aislamiento de ácidos nucleicos
NBT/BCIP	11681451001	Roche	Isoelectroenfoque
Neuraminidasa	11585886001	Roche	Isoelectroenfoque
N-glicosidasa F (PNGasa)	11365185001	Roche	Glicanos MALDI-TOF/MS

Reactivo / Producto	Referencia	Casa comercial	Aplicación
Nitrocelulosa 0,2µm	10600001	GE Healthcare Life Sciences	Isoelectroenfoque
Novex Sharp Pre-Stained Protein Standards	LC5800	Invitrogen	Western Blot
NP40 (Nonidet 40)	N6507	Sigma	Glicanos MALDI-TOF/MS
NuPAGE LDS 4x sample buffer	NP0008	Invitrogen	Western Blot
NuPAGE MES SDS Running buffer	NP0002	Invitrogen	Western Blot
OptiMEM I Reduced Serum Medium (1x) with glutamine	31985-047	Invitrogen	Transfección
Películas de rayos X RP2 PLUS	ENKLT	Agfa	Inmunodetección
Phast Gel Dry IEF	17-0677-01	GE Healthcare Life Sciences	Isoelectroenfoque
Phast Gel IEF 4-6,5	17-0544-01	GE Healthcare Life Sciences	Isoelectroenfoque
Prolong	P-36970	ThermoFischer	Inmunofluorescencia
Proteinase K	03115879001	Roche	Aislamiento de ácidos nucleicos
QIAquick Gel Extraction kit	28704	Qiagen	Purificación de banda
Quanti-iT™ PicoGreen® ds DNA Assay kit	P7589	Life Technologies	Cuantificación de ácidos nucleicos
SDS (Sodio dodecilsulfato)	L4390	Sigma	Glicanos MALDI-TOF/MS
SuperScript III First Strand Synthesis System	18080-051	Invitrogen	Reversotranscripción de RNA
SuperScript VILO	11754-050	Invitrogen	Reversotranscripción de RNA
SureSelect XT Human All Exon 50Mb	5190-4626	Agilent	WES
SureSelect XT Human All Exon V4+UTRs	5190-4636	Agilent	WES
Trifluoroacético (TFA)	34957	Sigma	Glicanos MALDI-TOF/MS
Tripure Isolation Reagent	15596-018	Invitrogen	Aislamiento de ácidos nucleicos
Tritón X100	11332481001	Roche	Inmunofluorescencia
TruSeqExome	FC-121-1008	Illumina	WES
TruSight™ One Sequencing Panel	FC-141-1006	Illumina	CES
Tween-20	822184	Merck	Isoelectroenfoque y WB
Urea	8487	Merck	Isoelectroenfoque
Yodometano	289566	Sigma	Glicanos MALDI-TOF/MS

## 1.3 Anticuerpos

Tabla 4. Anticuerpos.

Anticuerpo	Proteína - epítipo	Especie / descripción	Aplicación / dilución	Referencia	Casa comercial
<b>Anticuerpos primarios</b>					
Anti-ApoCIII	Apolipoproteína CIII	Conejo, pAb	IEF, 1/1000	PAB1285	Abnova
Anti- $\alpha$ -tubulina	Tubulina	Ratón, mAb	WB, 1/5000	T9026	Sigma-Aldrich
Anti-calnexina	Calnexina	Conejo, pAb	IF, 1/200	SPC-108	StressMarq
Anti-CCDC115	CCDC115	Conejo, pAb	IF, 1/50	20636-a-AP	Proteintech
Anti-CD63	CD63	Ratón, mAb	IF, 1/100	H5C6	DSHB
Anti-COG7	COG7	Ratón, mAb	WB, 1/1000	H00091949-M01	Abnova
Anti-EEA1	EEA1	Ratón, mAb	IF, 1/200	610457	BD Biosciences
Anti-GM130	GM130	Ratón, mAb	IF, 1/350	610823	BD Biosciences
Anti-ICAM1	ICAM-1	Conejo, pAb	WB, 1/1000	Sc-7891	Santa Cruz Biotechnology
Anti-ManII	Manosidasa II	Conejo, pAb	IF, 1/100	AB3712	Merck-Millipore
Anti-Transferrin	Transferrina	Conejo, pAb	IEF, 1/2000	194-5355	BioRad
<b>Anticuerpos secundarios</b>					
Anti-conejo IgG-Alexa488	IgG conejo	Burro	IF, 1/500	A-21206	ThermoFischer
Anti-conejo IgG Alexa555	IgG conejo	Burro	IF, 1/500	A-31572	ThermoFischer
Anti-conejo IgG-AP	IgG conejo	Cabra	IEF, 1/2000	A3687	Sigma
Anti-conejo IgG-HRP	IgG conejo	Cabra	WB, 1/5000	Sc-2030	Santa Cruz Biotechnology
Anti-ratón IgG-Alexa488	IgG ratón	Burro	IF, 1/500	A-21202	ThermoFischer
Anti-ratón IgG-Alexa594	IgG ratón	Burro	IF, 1/500	A-21203	ThermoFischer
Anti-ratón IgG-Alexa647	IgG ratón	Burro	IF, 1/500	A-31571	ThermoFischer
Anti-ratón IgG-HRP	IgG ratón	Cabra	WB, 1/10000	Sc-2005	Santa Cruz Biotechnology

mAb: Anticuerpo monoclonal; pAb: anticuerpo policlonal.

## 1.4 Oligonucleótidos

Tabla 5. Oligonucleótidos.

Gen	Nombre	Secuencia 5'-3'	Tm (°C)	Uso
AGL	AGL-3S	AAACAGCATTAGGTTTGCGG	54	PCR
AGL	AGL-3AS	TGCACTGCCATGGTTCATAC	56	PCR
AGL	AGL-4S	CATTTTATTTGGGACAATTAACCTTT	51	PCR
AGL	AGL-4AS	TTCTTTACTGCAATGAGAGAATGG	53	PCR
AGL	AGL-5S	TGAACCCAAGTGTTTGACCTC	55	PCR
AGL	AGL-5AS	ATGTTAGCCCCCTTTCCTTTC	52	PCR
AGL	AGL-15/16S	CACTTGGTGACAATCATACCC	53	PCR
AGL	AGL-15/16AS	CTTGGGATTCCTGAAAGCAG	54	PCR
AGL	AGL-20S	TTTGAAAACCTTGCTCCAGGAAG	53	PCR
AGL	AGL-20AS	GGTCCTTGACCGTACTTTG	56	PCR
AGL	AGL-24S	TCTGTGGACTGTAAGCCAGC	57	PCR
AGL	AGL-24AS	CACACATAAATGCCTGGTGC	55	PCR
AGL	ALG-26S	GGGAGTCACACAGATTGTTAGC	56	PCR
AGL	AGL-26AS	GTACCAGCAGGTGCCAAATC	57	PCR
AGL	AGL-31S	GCATCTACACTCAAATTCTGACC	54	PCR
AGL	AGL-31AS	CATCTCCTTTTGTTGCCCTG	54	PCR
AGL	AGL-32/33S	TGTACTAATGCCGAGCTTATTCTG	55	PCR
AGL	AGL-32/33AS	AGGCTTCACTTTTCATTGGG	53	PCR
AGL	AGL-1F	GCTCAAAGGCAGGAAACAAC	55	RT-PCR
AGL	AGL-1R	CACATTTCCCAAAGGCAAT	52	RT-PCR
AGL	AGL-2F	GCCAGCAGCAATTGAAGAAT	54	RT-PCR
AGL	AGL-2R	GTGGCAACTCCAGCTTGTTT	56	RT-PCR
AGL	AGL-3F	AGGAAGTGCCTCAAATGTGC	56	RT-PCR
AGL	AGL-3R	TGTATGCCCTGCATGTGTTT	55	RT-PCR
AGL	AGL-4F	TTCTCAAGTGCCAGTTTCC	55	RT-PCR
AGL	AGL-4R	AAACAACCTGAGCCTGTGCAT	55	RT-PCR
ALDOB	ALDOB-5S	CAATCCTCAAGCACAGTGGA	55	PCR
ALDOB	ALDOB-5AS	CTGGTGAGGGAAAAGGAGGT	57	PCR
ALG1	ALG1-E7-S	TCTCCTGGGTCCCGGGCCTGCTCTA	80	PCR
ALG1	ALG1-E7-AS	GTTCTCATTGAGACCGTGGCAGGGTGGA	81	PCR
ALG1	ALG1-E12-S	ATGGGGTGTGTGGGGGAGCCCAGG	81	PCR
ALG1	ALG1-E12-AS	CAACCGCAGCCCGGCCACCCA	85	PCR
ALG6	ALG6-12S	TGCATACTAGTGTTGTTTGTATTTT	51	PCR
ALG6	ALG6-12AS	CATTTGTGTAGTTTTGTTTGCATT	52	PCR
ANO5	ANO5-E5-S	TCCTCCTGAGTCATTATTGCTT	62	PCR
ANO5	ANO5-E5-AAS	GGGCCTCTCACTTTTACAACA	63	PCR
B4GALT1	B4GALT1-2S	ATCTGATGCAGCCTGTGGAG	61	PCR
B4GALT1	B4GALT1-2AS	TGTGAAATCACTCCCCTTCC	59	PCR
CCDC115	CCDC115-E1aI3-S	GAGGATCTCGTGTGCCTATTG	55	PCR
CCDC115	CCDC115-E1aI3-aS	GGGATGCTCTCTTATCGC	55	PCR

Gen	Nombre	Secuencia 5'-3'	Tm (°C)	Uso
COG7	COG7-E2-S	ATTTTACTGGTGGCTGGTCC	55	RT-PCR
COG7	COG7-E2-AS	TGCTGTTCTAAAGCCCTTACG	55	RT-PCR
COG7	COG7-F1-F	TATGGTAACGGCCAACGACT	56	RT-PCR
COG7	COG7-F1-R	GTCAATCACTTCCCCATGCT	55	RT-PCR
COG7	COG7-F2-F	CCAGGCTGCTGGAGTTCTAC	58	RT-PCR
COG7	COG7-F2-R	CCTGGCTTCATGATCCATCT	54	RT-PCR
CPT2	CPT2-3S	CTGTTGGGGACCCAAAACCTC	56	PCR
CPT2	CPT2-3AS	CCAAAACAGGAGAGAATATGCC	54	PCR
DOLK	HDK1-1F	GAACGGAGGGGAGAAGGTTG	55	PCR
DOLK	HDK1-1R	CCGCTTGGCATTCTGGTAC	56	PCR
DPAGT1	DPAGT1-3S	AGAACTGGGGTTATTTGGGC	60	PCR
DPAGT1	DPAGT1-3AS	CCAGAAAGGATCTGGGAAGAC	60	PCR
DPAGT1	DPAGT1-6S	GGCTAGTCCAGTGCCCATAG	60	PCR
DPAGT1	DPAGT1-6AS	GCCCACTCCCTTTGCAC	60	PCR
DPM1	DPM1-8S	GCTTTATACATACCAGTGAACCTCTTG	54	PCR
DPM1	DPM1-8AS	ACTAGAAGACCTCCACAATTTAGC	54	PCR
DPM1	DPM1-9S	CTTTAAGGACCTCTGTTGTAGGAC	55	PCR
DPM1	DPM1-9AS	AAAGATGCATGAAATTTACCTTACC	52	PCR
DPM1	DPM1-F	GTCCGACCCGTCCTTTTC	56	RT-PCR
DPM1	DPM1-R	GGTGGTCTTCATAAAAAGATGC	52	RT-PCR
EXT1	EXT1-1.1S	GTCTTTACAGGCGGGAAG	53	PCR
EXT1	EXT1-1.1AS	GACTCTGCACTTTGGATCTC	53	PCR
FOLR1	FOLR1-3S	GGCTGGGGAATCAAGGACTA	56	PCR
FOLR1	FOLR1-3AS	TAGAACTGGAGCTGGGGTTG	57	PCR
FOLR1	FOLR1-5S	GCAGGAATGGAGGGATTTGG	56	PCR
FOLR1	FOLR1-5AS	AATAGTTGGGAGCTGTGGGG	57	PCR
GAA	GAA-2.1S	GTTGATGTCTCAGAGCTGC	53	PCR
GAA	GAA-2.1AS	TGCTTTGCAGGGATGTAG	52	PCR
GAA	GAA-2.2S	GAGCAGTGCCACACAG	56	PCR
GAA	GAA-2.2AS	GTCTGCTCACACCTCCG	55	PCR
GALT	GALT-4-5S	AGGTAACCTGGCTCCAAC	54	PCR
GALT	GALT-4-5AS	CAGAACCAAAGCTTCATCAC	52	PCR
GALT	GALT-6-7S	GTGATGAAGCTTTGGTTCTG	52	PCR
GALT	GALT-6-7AS	CATGGGATAAGAAAGTTAGGG	50	PCR
GFPT1	GFPT1-E3-S	GCTATCCTGGTTCTTGACACAC	63	PCR
GFPT1	GFPT1-E3-AS	CCTCCCTTCAATGTTACAGACC	64	PCR
LIPA	LIPA-4S	GAGTACATCACTATGTCAATCTTTC	51	PCR
LIPA	LIIPA4-AS	CCCTCTCATACAACTTCAGAG	52	PCR
LIPA	LIPA-8S	TTTTAGTGCTTTGAAGGGC	51	PCR
LIPA	LIPA-8AS	TCAGATTTGTAAGCAGGTTG	50	PCR
NKX2-5	NKX2-51-S	CACCATGCAGGGAAGCTG	66	PCR
NKX2-5	NKX2-5-E1-AS	CGTGTCTCCTCCTCCTGG	64	PCR

Gen	Nombre	Secuencia 5'-3'	Tm (°C)	Uso
PHKA2	PHKA2-12S	AATTGGCCTGGAGGATGAG	55	PCR
PHKA2	PHKA2-12AS	CACGCCTGGCAGAGAGTC	58	PCR
PHKA2	PHKA2-14S	ATGTCACCAGGCAGAAGAGG	57	PCR
PHKA2	PHKA2-14AS	GTGGAACCGCCTGCTTTAG	58	PCR
PHKA2	PHKA2-15S	GGCAGTAGAAGAACCAAGCC	56	PCR
PHKA2	PHKA2-15AS	TGTACTAGGCTCTTGGTATCATTTTC	54	PCR
PHKA2	PHKA2-25S	CTGGTGTGGGAGGGGTC	58	PCR
PHKA2	PHKA2-25AS	CCTAAGCAGTTCTTCCCCTG	55	PCR
PHKA2	PHKA2-26S	CTGCTATCACCTCCCTCCC	57	PCR
PHKA2	PHKA2-26AS	TCTTCCACATTCGCTATGC	55	PCR
PHKA2	PHKA2-32S	CCAGGCACTGCTACGGTC	58	PCR
PHKA2	PHKA2-32AS	GATGACATTTTCTTGGTTCGAG	52	PCR
PHKB	PHKB-4S	TCACCAGAACACAGGATTAGC	54	PCR
PHKB	PHKB-4AS	CCTAGGATGTGGCTTTCCC	55	PCR
PHKB	PHKB-5S	GGTCCTCTGCCTTAGCTTTG	56	PCR
PHKB	PHKB-5AS	TCATCCCAAAACCATCTAGC	52	PCR
PMM2	PMM2-1S	GACGTGCCCTGCGACTC	59	PCR
PMM2	PMM2-1AS	GTGGTCGATAGCCCCAACTG	58	PCR
PMM2	PMM2-2S	TTGTGTTACCCTTAGAGTTTTGG	53	PCR
PMM2	PMM2-2AS	GGGCAGCCTATGATACTTGG	55	PCR
PMM2	PMM2-5S	TTGACCACACTAGCCTCTGC	57	PCR
PMM2	PMM2-5AS	TTTAACTACTGGCCACAAAGG	53	PCR
PMM2	PMM2-7S	AACCTTGGCAACCCACTAAC	55	PCR
PMM2	PMM2-7AS	GCGGGTGGGTAAAGTGTAGG	58	PCR
PYGL	PYGL-14/15S	AGACCACCCAATTCTTC	50	PCR
PYGL	PYGL-14/15AS	GTCAAACGCATTGACAG	50	PCR
RFT1	RFT1-E3-S	TATCTCCTCCCTGTTGCTGG	56	PCR
RFT1	RFT1-E3-AS	GCAATTCAGCTTTAGGCACG	55	PCR
SLC37A4	SLC37A4-3S	CTCCTTCCGGGACCTCCT	59	PCR
SLC37A4	SLC37A4-3AS	CACCACATGCCTGTACACAA	55	PCR
SLC37A4	SLC37A4-10S	CCACCTGTGCTTTGCATATAGA	55	PCR
SLC37A4	SLC37A4-10AS	TGTTTGTCTGAAGGCAGGAA	54	PCR
SRD5A3	SRD5A3-1S	AGATTAGGGACTGCGAGC	54	PCR
SRD5A3	SRD5A3-1AS	CTCCAGCTGAGTGCAAG	59	PCR
SSR4	SSR4-E3-S	CAGCTGATCCGTGCTATGAG	64	PCR
SSR4	SSR4-E3-AS	CCTCTATCGACACAGGTCCC	64	PCR

## 1.5 Equipos

**Tabla 6. Equipos.**

Equipo	Casa comercial	Aplicación
Bioanalyzer 2100	Agilent	Secuenciación masiva
HiSeq2000	Illumina	Secuenciación masiva
HPLC-Variant	BioRad	%CDT
iBlot® 7-Minut Blotting System	Invitrogen	Inmunotransferencia
MagNA Pure Compact	Roche	Aislamiento de ácidos nucleicos
Microscopio de barrido láser confocal LSM510 Meta	Zeiss	Estudios de colocalización
Microscopio electrónico de transmisión JEM1010	Jeol	Análisis ultraestructural del RE
Microscopio invertido Axiovert200	Zeiss	Inmunofluorescencia
MiSeq	Illumina	Secuenciación masiva
Nanodrop ND-1000	Thermo Scientific	Cuantificación de ácidos nucleicos
NextSeq500	Illumina	Secuenciación masiva
NuPAGE <i>Electrophoresis System</i>	Invitrogen	Western Blot
<i>Phast System</i>	GE Healthcare Life Sciences	Isoelectroenfoque
Qubit® 2.0 Fluorometer	Invitrogen	Cuantificación de ácidos nucleicos
SpeedVac	Savant	Concentración de ácidos nucleicos
SOLiD® 4	Life Technologies / APG	Secuenciación masiva
Veriti® Thermal Cycler	Applied Biosystems	Amplificación de ácidos nucleicos

Algunos equipos de uso habitual en el laboratorio no se especifican en la tabla.

## 1.6 Soporte informático

Tabla 7. Programas y bases de datos.

Programas		
Nombre	Aplicación	URL
Annovar	Anotación de variantes de datos de secuenciación masiva	<a href="http://annovar.openbioinformatics.org">http://annovar.openbioinformatics.org</a>
AST (Analyzer Splice Tool)	Predicción de sitios conservados de <i>splicing</i>	<a href="http://ibis.tau.ac.il/ssat/SpliceSiteFrame.htm">http://ibis.tau.ac.il/ssat/SpliceSiteFrame.htm</a>
BDGP (Berkeley Drosophila Genome Project)	Predicción de sitios conservados de <i>splicing</i>	<a href="http://www.fruitfly.org/seq_tools/splice.html">http://www.fruitfly.org/seq_tools/splice.html</a>
BedTools	Análisis de cobertura de secuenciación masiva	<a href="http://bedtools.readthedocs.org">http://bedtools.readthedocs.org</a>
BioMol	Análisis de datos de secuenciación masiva	<a href="http://www.biomol-informatics.com/Biomol-NGS">http://www.biomol-informatics.com/Biomol-NGS</a>
BLAST	Alineamiento de secuencias	<a href="http://blast.ncbi.nlm.nih.gov/Blast.cgi">http://blast.ncbi.nlm.nih.gov/Blast.cgi</a>
Bowtie	Alineamiento de lecturas de secuenciación masiva	<a href="http://bowtie-bio.sourceforge.net">http://bowtie-bio.sourceforge.net</a>
BWA	Alineamiento de lecturas de secuenciación masiva	<a href="http://bio-bwa.sourceforge.net">http://bio-bwa.sourceforge.net</a>
Chromas	Análisis de secuencias nucleotídicas	<a href="http://technelysium.com.au">http://technelysium.com.au</a>
Clustal Omega	Alineamiento de secuencias proteicas	<a href="http://www.ebi.ac.uk/Tools/msa/clustalo">http://www.ebi.ac.uk/Tools/msa/clustalo</a>
Cryp Skip	Análisis de mutaciones de <i>splicing</i>	<a href="http://cryp-skip.img.cas.cz">http://cryp-skip.img.cas.cz</a>
Cutadapt	Eliminación de la secuencia del adaptador de las lecturas de secuenciación masiva	<a href="http://cutadapt.readthedocs.org">http://cutadapt.readthedocs.org</a>
DNA Nexus	Análisis de calidad, alineamiento, anotación y análisis de variantes de secuenciación masiva	<a href="https://www.dnanexus.com">https://www.dnanexus.com</a>
ESE Finder 3.0	Análisis de mutaciones de <i>splicing</i>	<a href="http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi">http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi</a>
Exon Primer	Diseño de oligonucleótidos	<a href="https://ihg.helmholtz-muenchen.de/ihg/ExonPrimer.html">https://ihg.helmholtz-muenchen.de/ihg/ExonPrimer.html</a>
FastQC	Análisis de calidad de las lecturas de secuenciación masiva	<a href="http://www.bioinformatics.babraham.ac.uk/projects/fastqc">http://www.bioinformatics.babraham.ac.uk/projects/fastqc</a>
Fiji v1.0	Procesamiento y análisis de imágenes de microscopía	<a href="http://fiji.sc/Fiji">http://fiji.sc/Fiji</a>
GATK (Genome Analysis Toolkit)	Alineamiento de lecturas y detección de variantes de secuenciación masiva	<a href="https://www.broadinstitute.org/gatk">https://www.broadinstitute.org/gatk</a>
Genegrid (Genomatix)	Análisis de variantes de secuenciación masiva	<a href="https://www.genomatix.de/solutions/genegrid.html">https://www.genomatix.de/solutions/genegrid.html</a>
GlycoMod	Predicción de la masa molecular de oligosacáridos	<a href="http://web.expasy.org/glycomod">http://web.expasy.org/glycomod</a>



Nombre	Aplicación	URL
HSF 3.0 (Human Splicing Finder)	Análisis de mutaciones de <i>splicing</i>	<a href="http://www.umd.be/HSF3">http://www.umd.be/HSF3</a>
Huygens professional deconvolution software	Procesamiento y análisis de imágenes de microscopía	<a href="https://svi.nl/HuygensProfessional">https://svi.nl/HuygensProfessional</a>
IGV (Integrative Genomics Viewer)	Visor de lecturas mapeadas de secuenciación masiva	<a href="https://www.broadinstitute.org/igv">https://www.broadinstitute.org/igv</a>
miRror Suite	Predicción de miRNAs que se unirán a una secuencia	<a href="http://www.proto.cs.huji.ac.il/mirror">http://www.proto.cs.huji.ac.il/mirror</a>
Mutalyzer	Comprobación de la nomenclatura de mutaciones	<a href="https://mutalyzer.nl/">https://mutalyzer.nl/</a>
Mutation Taster	Predicción de la patogenicidad de cambios aminoacídicos	<a href="http://www.mutationtaster.org">http://www.mutationtaster.org</a>
Picard Tools	Análisis de alineamientos de secuenciación masiva	<a href="http://broadinstitute.github.io/picard">http://broadinstitute.github.io/picard</a>
PITA	Predicción de miRNAs que se unirán a una secuencia	<a href="http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html">http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html</a>
Polyphen 2	Predicción de la patogenicidad de cambios aminoacídicos	<a href="http://genetics.bwh.harvard.edu/pph2">http://genetics.bwh.harvard.edu/pph2</a>
Primer 3	Diseño de oligonucleótidos	<a href="http://bioinfo.ut.ee/primer3-0.4.0">http://bioinfo.ut.ee/primer3-0.4.0</a>
Pymol	Modelado de estructuras proteicas	<a href="http://www.pymol.org">http://www.pymol.org</a>
Quantity One (BioRad)	Cuantificación de bandas de WB	<a href="http://www.bio-rad.com/es-es/product/quantity-one-1-d-analysis-software">http://www.bio-rad.com/es-es/product/quantity-one-1-d-analysis-software</a>
RepeatMasker	Búsqueda de secuencias repetidas de DNA	<a href="http://www.repeatmasker.org">http://www.repeatmasker.org</a>
SAMtools	Análisis de alineamientos de secuenciación masiva	<a href="http://samtools.sourceforge.net">http://samtools.sourceforge.net</a>
Sequencher 5.2.4	Análisis de secuencias nucleotídicas	<a href="https://www.genecodes.com">https://www.genecodes.com</a>
SIFT	Predicción de la patogenicidad de cambios aminoacídicos	<a href="http://sift.jcvi.org">http://sift.jcvi.org</a>
SignalIP	Predicción de péptidos señal en secuencias aminoacídicas	<a href="http://www.cbs.dtu.dk/services/SignalP">http://www.cbs.dtu.dk/services/SignalP</a>
SNPCheck 3	Comprobación del lugar de hibridación de oligonucleótidos	<a href="https://secure.ngri.org.uk/SNPCheck/snpcheck.htm">https://secure.ngri.org.uk/SNPCheck/snpcheck.htm</a>
SPSS	Análisis estadístico	<a href="http://www-01.ibm.com/software/es/analytics/spss">http://www-01.ibm.com/software/es/analytics/spss</a>
STRING	Interacciones proteína-proteína	<a href="http://string-db.org">http://string-db.org</a>
SureDesign (Agilent)	Diseño de sondas para paneles personalizados de secuenciación masiva	<a href="https://earray.chem.agilent.com/suredesign">https://earray.chem.agilent.com/suredesign</a>
Variant Studio (Illumina)	Análisis de variantes de secuenciación masiva	<a href="https://www.illumina.com/informatics/research/biological-data-interpretation/variantstudio.html">https://www.illumina.com/informatics/research/biological-data-interpretation/variantstudio.html</a>

Bases de datos		
Nombre	Aplicación	URL
1000g (1000 Genomes Project Database)	Frecuencia alélica	<a href="http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes">http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes</a>
BioCarta Pathways	Rutas metabólicas y de señalización	<a href="http://www.genecarta.com">http://www.genecarta.com</a>
BioGPS	Expresión tisular de proteínas humanas	<a href="http://biogps.org/">http://biogps.org/</a>
dbSNP	Variaciones pequeñas en el genoma reportadas	<a href="http://www.ncbi.nlm.nih.gov/SNP">http://www.ncbi.nlm.nih.gov/SNP</a>
Ensembl	Genomas	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
Enzyme	Nomenclatura de las enzimas	<a href="http://enzyme.expasy.org">http://enzyme.expasy.org</a>
EVS (Exome Variant Server)	Frecuencia alélica	<a href="http://evs.gs.washington.edu/EVS">http://evs.gs.washington.edu/EVS</a>
ExAc (Exome Aggregation Consortium)	Frecuencia alélica	<a href="http://exac.broadinstitute.org">http://exac.broadinstitute.org</a>
GeneCards	Base de datos integrativa	<a href="http://www.genecards.org">http://www.genecards.org</a>
GeneHub GEPIS	Expresión tisular de proteínas humanas	<a href="http://research-public.gene.com/Research/genentech/genehub-gepis/">http://research-public.gene.com/Research/genentech/genehub-gepis/</a>
HGMD (Human Gene Mutation Database)	Mutaciones reportadas en humanos	<a href="http://www.hgmd.cf.ac.uk">http://www.hgmd.cf.ac.uk</a>
HGNC (HUGO Gene Nomenclature Committee)	Nomenclatura estandarizada de genes	<a href="http://www.genenames.org">http://www.genenames.org</a>
HGVS (Human Genome Variation Society)	Guía para la nomenclatura de mutaciones	<a href="http://www.hgvs.org">http://www.hgvs.org</a>
Human Protein Atlas	Expresión tisular de proteínas humanas	<a href="http://www.proteinatlas.org">http://www.proteinatlas.org</a>
Kegg Pathway (Kyoto Encyclopedia of Genes and Genomes)	Rutas metabólicas y de señalización	<a href="http://www.genome.jp/kegg/pathway.html">http://www.genome.jp/kegg/pathway.html</a>
NCBI (National Center for Biotechnology Information)	Genomas	<a href="http://ncbi.nih.gov">http://ncbi.nih.gov</a>
OMIM (Online Mendelian Inheritance in Man)	Fenotipos de enfermedades genéticas	<a href="http://www.omim.org">http://www.omim.org</a>
PDB (Protein Data Bank)	Proteínas	<a href="http://www.rcsb.org/pdb">http://www.rcsb.org/pdb</a>
UCSC (University of California, Santa Cruz)	Genomas	<a href="https://genome.ucsc.edu">https://genome.ucsc.edu</a>
UniProt	Proteínas	<a href="http://www.uniprot.org">http://www.uniprot.org</a>

## 2 MÉTODOS

---

### 2.1 Análisis bioquímico

#### 2.1.1 Cuantificación de las isoformas de transferrina por HPLC

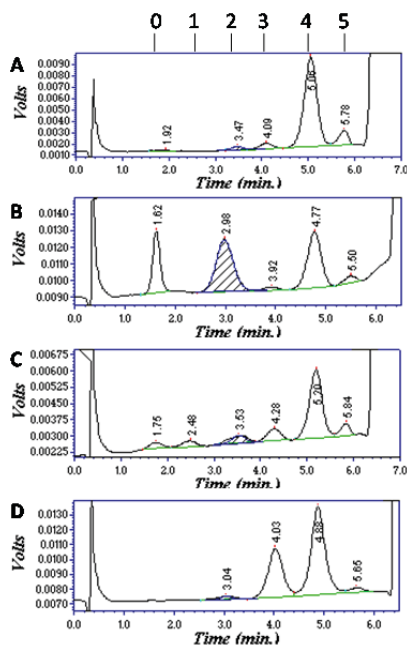
Se define como transferrina deficiente en carbohidratos al conjunto formado por las isoformas di-, mono- y a-sialiladas de la proteína. Se mide el porcentaje de transferrina deficiente en carbohidratos (%CDT) como método de cribado de CDG en los plasmas/sueros de pacientes con sospecha de enfermedad genética que presenten una clínica multisistémica con afectación neurológica de origen inexplicable referidos al Centro de Diagnóstico de Enfermedades Moleculares (CEDEM).

La cuantificación de las isoformas de transferrina se realizó con el kit Ready-Prep %CDT by HPLC™ siguiendo las indicaciones del proveedor y se analizó en un HPLC-Variant, que contiene una columna de intercambio aniónico y un detector UV. Cada isoforma se identifica por su tiempo de retención y el área del pico es proporcional a su cuantía. Los resultados se expresan como el porcentaje de cada isoforma respecto de la suma de a-, mono-, di-, tri-, tetra- y penta-sialotransferrina.

A partir del perfil del cromatograma y de los valores de cada isoforma de la transferrina comparado con los valores control de población pediátrica o adulta según el caso, se clasificaron los pacientes con:

- Patrón normal: Se identifica la isoforma mayoritaria, la tetrasialotransferrina, además de tri- y pentasialotransferrina (**Figura 5 A**).
- Patrón tipo 1: Se caracteriza por un aumento de las isoformas di- y asialiladas de la transferrina y una disminución de la tetrasialilada (**Figura 5 B**). Este patrón es característico de pacientes CDG-I, pero también de la galactosemia y la intolerancia a la fructosa (HFI).
- Patrón tipo 2: caracterizado por un aumento variable de cualquiera de las isoformas hipoglicosiladas de la proteína (**Figura 5 C y D**). Este perfil es típico de los pacientes CDG-II y de pacientes con daño hepático.

Las muestras de plasma con EDTA presentan un pico que coeluye con las formas a- y monosialiladas e imposibilita la determinación del %CDT. Tampoco son valorables las muestras hemolizadas o con altos niveles de bilirrubina por una subida de la línea base. Además, las muestras de pacientes con anemia pueden tener un área total de transferrina tan bajo que dificulte la determinación de los niveles del %CDT.



**Fig 5. Medida del %CDT mediante HPLC de (A)** un control, donde se aprecia que la isoforma mayoritaria es la tetrasialilada, **(B)** un paciente CDG-I que presenta un patrón típico tipo 1, caracterizado por aumento en las isoformas di- y asialiladas, **(C)** un paciente CDG-II con un patrón tipo 2 debido a un aumento de todas las isoformas de bajo contenido en ácido siálico y **(D)** otro paciente CDG-II con otro patrón tipo 2 distinto caracterizado por un aumento de la isoforma trisialilada de la transferrina. Los números de la parte superior indican el número de ácidos siálicos que tiene la proteína. En el eje de las X se representa el tiempo de retención de cada isoforma en minutos.

2.1.2 Isoelectroenfoque de transferrina, transferencia y revelado

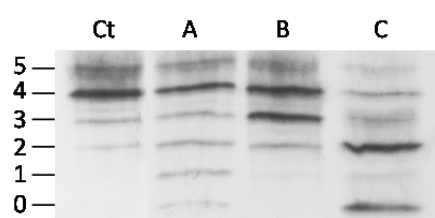
El isoelectroenfoque permite la visualización de las isoformas de transferrina de manera específica eliminándose los problemas de interferencia que ocurren en la detección por HPLC. El isoelectroenfoque de la proteína N-glicosilada transferrina (Trf-IEF) se realizó tal y como describe (Hackler et al., 1995). Brevemente, 3µL de plasma se saturaron en Fe<sup>+3</sup> con una solución de citrato férrico 5mM en bicarbonato sódico 0,25M durante 30 minutos a temperatura ambiente. Se centrifugó y 5µL del sobrenadante se diluyeron 1/300 en agua. Las muestras se corrieron en un gel Phast Gel IEF 4-6,5 en un *Phast System* bajo las siguientes condiciones (**Tabla 8**).

Tabla 8. Etapas del isoelectroenfoque de transferrina.

	Paso	AVh	Aplicador	V	mA	W	T °C	Vh
Preenfocado	5.1			2000	2,0	3,5	15	75
Aplicación muestra	5.2	0	Down					
IEF	5.2			200	2,0	3,5	15	15
	5.3	0	Up					
	5.3			2000	5,0	3,5	15	200
Modo de espera	5.4			50	0,5	0,5	15	50

Una vez finalizado el isoelectroenfoque (IEF), las proteínas separadas por su punto isoeléctrico se transfirieron a una membrana de nitrocelulosa de 0,2µm por difusión pasiva durante 30 minutos. Posteriormente, la membrana se bloqueó con solución de bloqueo, leche al 5% en PBS Tween-20 0,05%, durante una hora a temperatura ambiente y se incubó con anticuerpo anti-transferrina humana diluido en solución de bloqueo durante dos horas a temperatura ambiente con agitación (ver condiciones de cada uno de los anticuerpos utilizados en **Tabla 4**). Tras tres lavados de 10 minutos con PBS Tween-20 0,05%, la membrana se incubó con anticuerpo anti-IgG de conejo unido a fosfatasa alcalina diluido en solución de bloqueo durante una hora a temperatura ambiente con agitación. Se lavó de nuevo tres veces con el tampón de lavado y se reveló con NBT/BCIP diluido 1/10 en una solución al 5% de Tris-HCl 1M pH 9,7 hasta la aparición de bandas. Los resultados se valoraron de forma cualitativa por comparación con un suero control y se analizó el patrón obtenido de la misma manera que por HPLC (**Figura 6**).

La presencia de polimorfismos en la propia secuencia aminoacídica de la Trf que pudiesen originar un perfil anómalo fue excluido por digestión con neuraminidasa, que elimina los residuos de ácido siálico, único azúcar responsable de la carga del glicano. Brevemente, el plasma se incubó con neuraminidasa 10U/mL (1:1) a 37°C O/N. Una vez finalizada la digestión, las muestras se saturaron en  $\text{Fe}^{+3}$  por incubación con citrato férrico 5mM durante 30 minutos a RT, se diluyeron 1/600 en agua y se analizaron por IEF. Los resultados se valoraron de forma cualitativa por comparación con un suero control.



**Fig 6. Distintos patrones de Trf detectables por IEF** de una muestra control (**Ct**), un paciente CDG-II con un aumento de todas las isoformas hiposialiladas de la Trf (**A**), otro CDG-II con una elevación de trisialotransferrina (**B**) y un CDG-I con un aumento de di- y asialotransferrina (**C**). Los números de la izquierda indican el número de residuos de ácido siálico unidos a la proteína.

### 2.1.3 Isoelectroenfoque de apolipoproteína CIII, transferencia y revelado

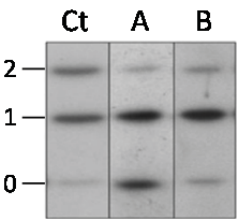
El IEF de la O-glicoproteína apolipoproteína CIII (ApoCIII) se realizó en todos los pacientes que presentaron un patrón tipo 2 en el Trf-IEF para valorar un posible defecto combinado en la ruta de la N- y O-glicosilación de proteínas. Está basado en el método descrito por (Wopereis et al., 2003). El *Phast Gel Dry IEF* se rehidrató durante una hora y media a temperatura ambiente con una solución de urea 8,5M y anfolinas 4,2-4,9 y 3,5-5,0 diluidos 1/25 y 1/50 respectivamente. 1µL de cada muestra se corrió en el gel rehidratado en un *Phast System* con las siguientes condiciones (**Tabla 9**).

Tabla 9. Etapas del isoelectroenfoque de apolipoproteína CIII.

	Paso	AVh	Aplicador	V	mA	W	T °C	Vh
Preenfocado	4.1			2000	2,0	5	10	75
Aplicación muestra	4.2	0	Down					
IEF	4.2			200	2,0	5	10	100
	4.3	0	Up					
	4.3			2000	2,0	5	10	495
Modo de espera	4.4			50	0,5	0,5	15	50

Las proteínas separadas en el gel según su punto isoeléctrico se transfirieron a una membrana de nitrocelulosa 0,2µm por difusión pasiva durante una hora. La membrana se bloqueó con solución de bloqueo durante una hora a temperatura ambiente y se incubó con anticuerpo anti-ApoCIII diluido en solución de bloqueo O/N a 4°C con agitación (ver condiciones en la **Tabla 4**). Se realizaron tres lavados con PBS Tween-20 0,05% y se incubó con anticuerpo anti-IgG de conejo unido a fosfatasa alcalina diluido en solución de bloqueo durante dos horas a temperatura ambiente con agitación. Se lavó de nuevo tres veces con el tampón de lavado y se reveló la membrana con NBT/BCIP diluido 1/10 en una solución al 5% de Tris-HCl 1M pH 9,7 hasta la aparición de bandas. Los resultados se valoraron de forma cualitativa por comparación con un suero control, observándose varios tipos de patrones de bandeo en el gel (Lefeber et al., 2011):

- Patrón normal: con di- y mono- en igual cuantía y trazas de asialo-ApoCIII (**Figura 7 Ct**).
- Patrón ApoCIII-0: con una disminución o incluso ausencia de di-, un ligero aumento de mono- y un gran aumento de asialo-ApoCIII (**Figura 7 A**). Este patrón es característico de los pacientes CDG debido a una deficiencia en el complejo oligomérico de Golgi (COG).
- Patrón ApoCIII-1: con una disminución de di- y un gran aumento de monosialo-ApoCIII. Asialo-ApoCIII está levemente aumentada (**Figura 7 B**). Es característico de los pacientes CDG debido a mutaciones en la subunidad α2 de la ATPasa tipo-V (*ATP6VOA2*).



**Fig 7.** Distintos patrones de ApoCIII detectables por IEF de una muestra control (**Ct**), un paciente CDG-II con una deficiencia en una subunidad del complejo COG (**A**) y un paciente CDG-II debido a un defecto en la subunidad α2 de la ATPasa tipo-V (**B**). Los números de la izquierda indican en el número de residuos de ácido siálico unidos a la proteína.

### 2.1.4 Análisis de N-glicanos de proteínas plasmáticas por MALDI-TOF/MS

El análisis de N-glicanos unidos a proteínas plasmáticas por MALDI-TOF acoplado a un espectrómetro de masas (MS) se realizó a pacientes con un patrón de Trf-IEF tipo 2. El perfil de N-glicanos liberados de las proteínas séricas del paciente puede sugerir la enzima deficiente en la ruta de procesamiento del glicano N-unido. El procedimiento está basado en el método publicado por (Morelle and Michalski, 2007).

Las proteínas plasmáticas se desnaturalizaron con sodio dodecilsulfo (SDS) 0,52% y beta-mercaptoetanol ( $\beta$ ME) 0,25% en bicarbonato amónico 50mM pH 8,3 a 100°C durante 20 minutos y nonidet 40 (NP40) 2,5% en bicarbonato amónico 50mM pH 8,3 durante 5 minutos a RT. Los glicanos N-unidos a las proteínas séricas se liberaron por digestión con 3U de N-glicosidasa F (PNGasa) a 37°C O/N. La muestra se evaporó a sequedad en corriente de nitrógeno y se redisolvió en 500 $\mu$ L de ácido trifluoroacético (TFA) 0,1%. Las proteínas precipitadas se eliminaron por centrifugación 5 minutos a 14.000rpm y con un filtro de 0,22 $\mu$ m. Los glicanos se purificaron en una columna de grafito de 150mg (Mills et al., 2003), se lavaron con agua y ACN 5% con TFA 0,1% y se eluyeron con 600 $\mu$ L de ACN 50% TFA 0,1%. La muestra se llevó a sequedad y se resuspendió en 500 $\mu$ L de dimetil sulfóxido (DMSO) con 30mg de hidróxido sódico en polvo en un baño de ultrasonidos durante 20 minutos. Los glicanos se permetilaron con 300 $\mu$ L de yodometano en atmósfera de nitrógeno durante 90 minutos y se purificaron por extracción líquido-líquido con 1mL de ácido acético 5% y 600 $\mu$ L de cloroformo. La fase orgánica se lavó con agua y se secó en corriente de nitrógeno. Los glicanos se resuspendieron en 500 $\mu$ L de metanol y se purificaron en una columna sep-pack C18. La columna se lavó con agua y ACN 10% y los glicanos se eluyeron con 3mL de ACN 80% y se llevaron a sequedad. La muestra se redisolvió en metanol:agua 1:1, se mezcló en una proporción 1:1 con 1 $\mu$ L de solución matriz DHB (metanol:acetato sódico 1mM 1:1 (v/v) y 10mg DHB/mL) y se llevó a un MALDI-TOF/MS (Ultraflex III, Bruker) para ser analizado en modo positivo.

Cada glicano se identificó por su masa molecular según el programa de acceso *on-line* Glycomod y se midió la intensidad del pico. Los resultados se expresaron como el porcentaje de cada uno de los glicanos frente a la suma total de glicanos. También se calculó el porcentaje de glicanos con tres antenas, con fucosa, truncados y con galactosa o N-acetilglucosamina terminal mediante la suma de los glicanos con esas características respecto del total de glicanos, y el porcentaje de glicanos sin ácido siálico terminal (nivel de hiposialización) como un cociente de la suma de las isoformas con un siálico terminal (m/z 2433 y su fucosilado, m/z 2607) entre la suma de las isoformas con dos siálicos terminales (m/z 2794 y su fucosilado, m/z 2968) (Guillard et al., 2011a).

## 2.2 Estudios celulares

### 2.2.1 Separación electroforética de proteínas en geles SDS-PAGE

Para la obtención de la fracción proteica de los fibroblastos en cultivo, estos se tripsinizaron, se lavaron con PBS y se centrifugaron durante 5 minutos a 1.400rpm. Se descartó el sobrenadante y el pellet se resuspendió en tampón de lisis (**Tabla 2**) con inhibidor de proteasas *mini EDTA-free protease inhibitor cocktail* y se agitó durante 20 minutos a 4°C. A continuación se centrifugó a 1.400rpm durante 5 minutos y se conservó el sobrenadante en hielo.

La cuantificación de la cantidad de proteína presente en el extracto total se llevó a cabo mediante el método de *Bradford*. Las proteínas contenidas en 20-50µg de extracto total se desnaturalizaron con tampón de carga *NuPAGE LDS Sample buffer 4x* por ebullición y se llevaron a un gel de electroforesis.

La separación electroforética se llevó a cabo en condiciones reductoras en los sistemas *NuPAGE* en geles *NuPAGE Bis-Tris* al 10% o geles *NuPAGE Bis-Tris* de gradiente del 4-12% y tampón de electroforesis MES (*2-(N-morpholino) ethanesulfonic acid*) 1x. Como marcador de peso molecular se usó *Novex Sharp Pre-Stained protein Standard*. Los geles se corrieron a voltaje constante (120V) durante 1,5-2h.

### 2.2.2 Electrotransferencia e inmunodetección de proteínas

Tras la electroforesis, los geles se transfirieron a una membrana de nitrocelulosa de 0,2µm con el sistema *iBlot® 7-Minute Blotting System* durante 7 minutos a 20V y se comprobó la eficiencia de la transferencia con Rojo Ponceau. Las membranas se bloquearon con solución de bloqueo (leche al 5% en PBS Tween-20 0,05%) durante 1h con agitación. Seguidamente se incubaron con el anticuerpo primario correspondiente diluido en solución de bloqueo (ver condiciones para cada uno de los anticuerpos utilizados en **Tabla 4**). Se realizaron tres lavados de 10 minutos con TBS Tween-20 0,05% y se incubaron con el anticuerpo secundario correspondiente unido a peroxidasa (HRP) durante 1h en agitación. Tras tres lavados con TBS Tween-20 0,05%, se realizó la detección por quimioluminiscencia con ECL y película de rayos X.

Para realizar la detección de otras proteínas sobre la misma membrana, estas se deshibridaron con solución de deshibridación (glicina 0,2M en TBS con Tween 20 0,05%) durante 20 minutos a 80°C. A continuación se bloquearon de nuevo y se continuó con el proceso de inmunodetección anteriormente descrito.

El escaneado de las películas y el densitometrado de las bandas obtenidas se realizó con el escáner BioRad G-8000 y el *software Quantity One* (BioRad).



### 2.2.3 Mantenimiento de los cultivos celulares

Los cultivos primarios de fibroblastos procedentes de biopsia de piel se cultivaron en botellas de 25 o 75cm<sup>2</sup> o placas estandarizadas de 6 pocillos M6 (Ø 35mm) o 24 pocillos M24 (Ø 16mm). Las células se crecieron a 37°C en una atmósfera de humedad relativa del 95% y 5%CO<sub>2</sub>, utilizando como medio de cultivo Medio Mínimo de Eagle (MEM) suplementado con glutamina 2mM, suero bovino fetal (FBS) al 10% (v/v), penicilina 100U/mL y estreptomicina 100µg/mL. Las células fueron recogidas con una solución de tripsina 0,25% y EDTA 0,02% y se sedimentaron mediante centrifugación. Las líneas celulares se conservaron en FBS con DMSO al 5% en nitrógeno líquido. En todos los ensayos las células se utilizaron al 70-80% de confluencia.

### 2.2.4 Detección de proteínas por inmunofluorescencia

Para la detección de proteínas en células se partió de 10<sup>5</sup> fibroblastos derivados de pacientes o del control (CC02509), que se sembraron en pocillos de M24 (Ø 16mm) sobre un cubreobjetos de 10mm de diámetro con 0,5mL de MEM completo suplementado con 10% FBS, glutamina y antibióticos. Las células se fijaron con formalina al 4% durante 20 minutos, se permeabilizaron con glicina 0,05mM en PBS durante 30 minutos y se lavaron tres veces con PBS. Posteriormente, se incubaron durante 30 minutos con solución de bloqueo (tritón X-100 0,1%, BSA 1%, FBS 20% en PBS) y seguidamente con los anticuerpos primarios correspondientes diluidos en solución de bloqueo durante una hora. Tras tres lavados con PBS, los cubreobjetos se incubaron durante una hora con la mezcla de los correspondientes anticuerpos secundarios unidos a un fluoróforo diluidos en solución de bloqueo (ver cada una de las condiciones de los anticuerpos utilizados en **Tabla 4**). Tras tres lavados con PBS y agua destilada, las células se deshidrataron con etanol y se montaron sobre los portaobjetos con Prolong. Las tinciones se visualizaron con un microscopio de fluorescencia. El análisis y procesamiento de las imágenes se realizaron con el *software* Fiji o con *Huygens professional deconvolution software*.

### 2.2.5 Evaluación del transporte retrógrado AG-RE

La evaluación del transporte retrógrado se realizó por tratamiento de las células con Brefeldin A (BFA), un metabolito fúngico que altera la estructura y función del AG induciendo la redistribución de las enzimas del Golgi, como Manosidasa II (Man II), al retículo endoplasmático vía transporte retrógrado. El método está basado en el protocolo publicado por (Reynders et al., 2009).

Tras 24 horas después de sembrar las células sobre los cubreobjetos en los pocillos de M24, se sustituyó el medio por MEM completo con 2,5µg de BFA y se incubó durante 6 minutos a 37°C. La reacción se detuvo eliminando el medio y fijando las células con formalina y se continuó con el protocolo de inmunofluorescencia descrito en el apartado 2.2.4. Como anticuerpos primarios se usaron anti-GM130, como proteína estructural del AG, y anti-Man II como enzima del AG. Las tinciones se visualizaron con un microscopio invertido Axiovert200 conectado a una cámara CCD RT Slider (Diagnostic) y controlado por el *software* Metavue 5.07 (Universal Imaging).

### 2.2.6 Estudios de colocalización

Con el fin de determinar la localización subcelular de la proteína CCDC115 se realizó una doble o triple tinción con su anticuerpo y los siguientes marcadores de orgánulos: calnexina para retículo endoplasmático, GM130 para aparato de Golgi, CD63 para lisosomas y EEA1 para endosomas tempranos, y se prosiguió con el protocolo de inmunofluorescencia detallado en el apartado 2.2.4. Las preparaciones se visualizaron con un microscopio de barrido láser confocal LSM510 Meta acoplado a un microscopio invertido Axiovert200 y controlado por el *software* ZEN 2008 sp2 (Zeiss). El análisis de las imágenes se realizó con los plugins RGB profiler y JACoP, ambos de Fiji.

### 2.2.7 Análisis ultraestructural del RE por microscopía electrónica

Para el análisis ultraestructural del RE se partió de una monocapa de células confluentes crecidas en una p60 (Ø 60mm) a la que se añadió fijador (4% paraformaldehído, 2% glutaraldehído en tampón fosfato sódico 0,1M pH 7,4) y se incubó durante 2 horas a RT. Posteriormente, se lavó tres veces con tampón sódico 0,1M pH 7,4. Las células se tiñeron con metales pesados (OsO<sub>4</sub> 1% y ferrocianuro potásico 1% 1 hora a 4°C, ácido tánico 0,15% en tampón fosfato sódico 0,1M pH 7,4 durante 1 minuto y acetato de uranilo 2% una hora a temperatura ambiente en oscuridad) y se deshidrataron con concentraciones crecientes de etanol (50, 75, 90, 95 y 100%) a 4°C. Seguidamente, se embebieron en una resina epoxi TAAB 812 (*TAAB laboratories*) durante 48 horas a 60°C. Los cortes de 70nm se realizaron con un ultramicrotomo Ultracut UCT (Leica) con una cuchilla de diamante (diatome) y se depositaron sobre una rejilla Cu/Pd recubierta con Fomvar y carbono. Las secciones se tiñeron con acetato de uranilo al 2% y citrato de plomo y se examinaron a 80Kv en un microscopio electrónico de transmisión JEM1010 acoplado a una cámara digital 4Kx4K TemCam-F416 (TVIPS).

## 2.3 Análisis genético

### 2.3.1 Aislamiento de ácidos nucleicos y cuantificación

Se utilizó el extractor automático *MagNA Pure Compact* siguiendo las recomendaciones del proveedor para cada uno de los kits empleados.

- Para la obtención de gDNA de muestras de sangre total se utilizó *MagNA Pure Compact DNA Isolation Kit-large volumen*.
- Para sangre en papel se utilizó *MagNA Pure LC DNA Isolation Kit-lysis y binding buffer refill* con proteinasa K 0,8ng/mL y *MagNA Pure Compact Nucleic Acid Isolation Kit I*.
- Cuando se partió de fibroblastos de piel cultivados se empleó *MagNA Pure Compact Nucleic Acid Isolation Kit I*.
- Para el aislamiento de RNA independientemente del tipo de muestra se utilizó *MagNA Pure Compact RNA Isolation Kit* o *Tripure Isolation Reagent*.

La cuantificación y determinación de la pureza de los ácidos nucleicos aislados se realizó en un Nanodrop. Para las muestras empleadas en secuenciación masiva se utilizó Quant-iT™ PicroGreen® como agente intercalante de dsDNA y Qubit® como medidor de fluorescencia.

### 2.3.2 Amplificación de ácidos nucleicos

Amplificación de gDNA: Se empleó el protocolo estándar (Mullis and Faloona, 1987), teniendo en cuenta la longitud del amplicón y la temperatura de disociación del par de oligonucleótidos específicos o T<sub>m</sub> (**Tabla 5**). Se amplificaron entre 250-500ng de gDNA utilizando la polimerasa *Fast Start TAQ* en un termociclador *Veriti® Thermal Cycler*.

Amplificación de cDNA: A partir de 1,5µg de RNA total se obtuvo cDNA por retrotranscripción utilizando *SuperScript III First Strand Synthesis System* y OligodT o *Superscript VILO* y *random hexamers*, siguiendo las recomendaciones del proveedor. Posteriormente el cDNA se amplificó mediante PCR y oligonucleótidos específicos (**Tabla 5**) y la polimerasa *Fast Start TAQ* en un termociclador *Veriti® Thermal Cycler*.

Los oligonucleótidos específicos se diseñaron con los programas de acceso *on-line Primer 3* o *Exon Primer* y se comprobó que no hibridaban en regiones donde pudiese haber polimorfismos descritos con *SNPCheck 3*. Tras la amplificación, los productos de PCR se separaron en un gel de agarosa del % adecuado según el tamaño del amplicón esperado y con bromuro de etidio (BrEt) al 4% (v/v).

### 2.3.3 Purificación de los productos de PCR y secuenciación por Sanger convencional

Tras la amplificación, los fragmentos de DNA se purificaron con *E.Z.N.A.® Cycle Pure Kit* o *ExoSAP-IT® PCR Product Cleanup*. Los fragmentos de DNA recortados de geles de agarosa se purificaron con el kit *QIAquick Gel Extraction kit*.

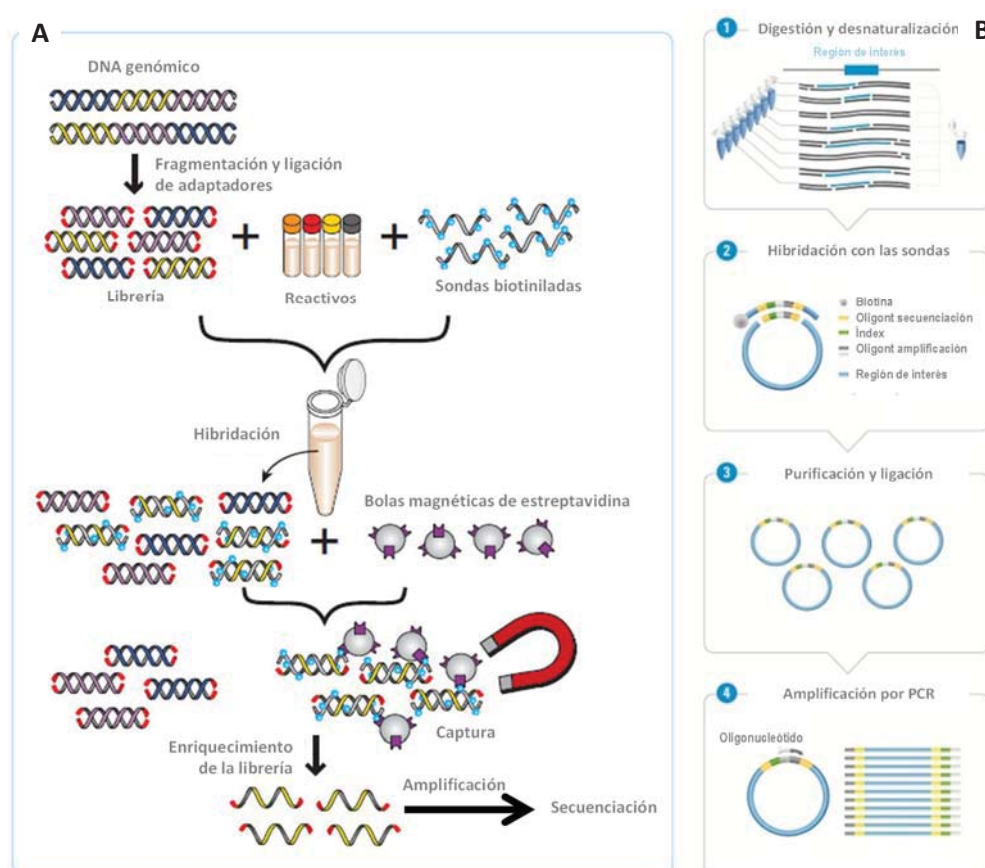
El DNA purificado se secuenció mediante el método de terminación de cadena de DNA por incorporación de dideoxinucleótidos trifosfato (ddNTPs) (Sanger et al., 1977). Se emplearon 40-50ng de DNA amplificado por PCR, 0,25µM de oligonucleótidos específicos y *BidDye® Terminator v3.1 Cycle Sequencing* y se analizaron con un secuenciador automático *ABI Prism® 3730xl* (Applied Biosystems). Las secuencias se analizaron con los programas Chromas o Sequencher 5.2.4.

### 2.3.4 Secuenciación masiva: Paneles de genes (TES)

Se han empleado dos paneles personalizados diferentes para el análisis de pacientes con sospecha de CDG y de glucogenosis.

Para el análisis de pacientes con sospecha de CDG se diseñó un panel de 43 genes asociados a CDG (TES-CDG) (**Anexo 3**) con el programa *SureDesign* y se siguió el procedimiento ilustrado en la **Figura 8A**. Brevemente, se partió de 1µg de DNA purificado con *MagNA Pure Compact* cuya concentración se midió por absorbancia a 260nm con un Nanodrop y su integridad se evaluó con un gel de agarosa 1%. Los DNA se fragmentaron por sonicación y se ligaron los adaptadores a ambos extremos. Seguidamente se hizo la captura de las regiones de interés con un kit personalizado de SureSelect que utiliza sondas biotiniladas que posteriormente se purifican con bolas de estreptavidina. Las librerías se amplificaron por PCR en emulsión y se secuenciaron en lecturas de 50pb en *paired-end* usando un SOLiD4®, que se basa en la secuenciación por ligación. La longitud total de secuencia fue de 16 x 0,2Mb. La calidad de las lecturas se analizó con FastQC 0.9.1, el alineamiento con Bowtie 0.12.7 y las lecturas duplicadas fueron detectadas y eliminadas con Picard Tools 1.48. El realineamiento y recalibrado de las lecturas así como la detección de variantes e indel se realizó con GATK y el análisis de cobertura con BedTools 2.12. El posterior análisis de variantes se realizó con la plataforma comercial *DNA Nexus classic* y con *BioMol Informatics*.

Para el estudio de pacientes con sospecha de glucogenosis, se diseñó mediante el *software SureDesign* un panel personalizado de Haloplex que captura 111 genes relacionados con enfermedades metabólicas hereditarias (TES-EMH), entre los que se incluyen 22 genes de glucogenosis (**Anexo 4**). El esquema de trabajo está ilustrado en la **Figura 8 B**. Brevemente, se partió de 300ng de DNA de pacientes con sospecha de glucogenosis purificado con *MagNA Pure Compact*. La concentración se midió mediante fluorimetría utilizando *PicoGreen* y la pureza del DNA se evaluó con un gel de agarosa del 1%. El DNA se digirió con enzimas de restricción y el resultado se analizó con un *Bioanalyzer*. Se capturaron los fragmentos con las secuencias de interés con sondas con biotina a la vez que se ligaron los adaptadores y se circularizaron. La librería se amplificó y los DNA se mezclaron. Los *cluster* se generaron por PCR en puente y se secuenciaron por síntesis en un MiSeq en lecturas de 150pb en *paired-end*. Los análisis de calidad de las lecturas así como el alineamiento y la anotación y el análisis de variantes se realizaron con la plataforma comercial *DNA Nexus*.



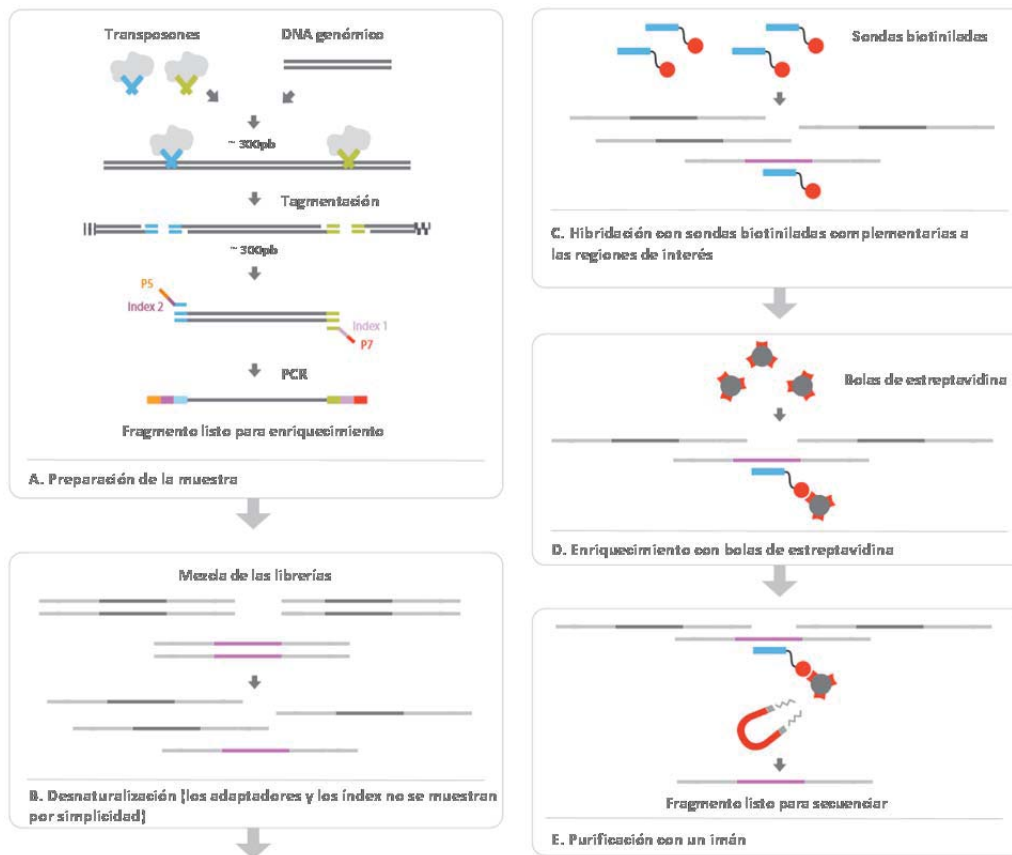
**Fig 8. Preparación de las muestras para TES.** (A) Esquema de trabajo utilizado en la preparación de la librería de TES-CDG con SureSelect. El gDNA se fragmenta y se unen unos adaptadores a ambos extremos de cada fragmento generado. La librería se enriquece con unas sondas biotiniladas y se purifican con avidina unida a bolas magnéticas. Finalmente, se secuencia la librería. (B) Flujo de trabajo con Haloplex empleado para la preparación de la librería de TES-EMH. El gDNA se digiere con enzimas de restricción. El enriquecimiento de la librería se realiza con unas sondas que tienen unida biotina y a la vez insertan el adaptador durante la amplificación. Figura adaptada de la página web de sus respectivas casas comerciales.

### 2.3.5 Secuenciación masiva: Exoma completo (WES)

La secuenciación del exoma completo se realizó en pacientes con sospecha de CDG utilizando sistemas de captura diferentes. El enriquecimiento de la librería se realizó con el kit *SureSelect XT Human All Exon 50Mb* (Agilent), o *SureSelect V4* (Agilent) o *TruSeqExome* (Illumina). La secuenciación se realizó en lecturas de 75pb o 100pb en *paired-end* en un HiSeq2000, que se basa en la secuenciación por síntesis. La calidad de las lecturas generadas con los kits de Agilent fue valorada con FastQC 0.9.1 y se eliminó la secuencia del adaptador con Cutadapt. El alineamiento se realizó con BWA. Las lecturas duplicadas se identificaron y descartaron con Picard Tools 1.48 y SAMTools y se reanalizó el alineamiento con GATK. La detección de variantes e indel se realizó con GATK y se anotaron con Annovar. El análisis de cobertura se realizó con BedTools 2.1.2 y la detección de variantes con *BioMol Informatics*. El procesamiento de las lecturas generadas con el kit de Illumina así como el alineamiento y la anotación de las variantes se realizó con los procesadores de datos comerciales de *DNA Nexus* y *Genomatix Genome Analyzer* y el análisis de variantes con las plataformas comerciales *Genomatix Genegrid* y *Variant Studio*.

### 2.3.6 Secuenciación masiva: Exoma clínico (CES)

El estudio del exoma clínico se realizó en pacientes con sospecha de CDG y de glucogenosis mediante el panel de genes *Illumina® Clinical Exome Sequencing (CES) TruSight™ One*, que permite la captura de 4.813 genes asociados a enfermedad y anotados en la base de datos de OMIM. En total se capturan 62.000 exones y 50pb de sus regiones intrónicas flanqueantes. El flujo de trabajo se esquematiza en la **Figura 9**. Se partió de 50ng de DNA que se purificó con *MagNA Pure Compact*, se cuantificó por fluorimetría con *PicoGreen*, y se comprobó su integridad en un gel de agarosa 1%. El DNA se fragmentó con transposones que a la vez insertan unos adaptadores (tagmentación) y el resultado se analizó en un gel de agarosa 1%. Posteriormente, mediante PCR se unieron otros adaptadores que contienen el índice y unas secuencias complementarias a los oligonucleótidos de amplificación y secuenciación. Los DNA de todos los pacientes se mezclaron y se realizó la captura de los fragmentos que contienen las regiones de interés con sondas biotiniladas y bolas magnéticas de estreptavidina. La librería se amplificó y se generaron los *cluster* por PCR en puente. La secuenciación por síntesis se realizó con un MiSeq o un NextSeq500 en lecturas de 250pb en *paired-end*. Los análisis de calidad de las lecturas así como el alineamiento y la anotación de variantes se realizaron con el procesador de datos comercial DNA Nexus. El análisis de variantes se realizó con las plataformas comerciales DNA Nexus y *Variant Studio*.



**Fig 9. Preparación de las muestras para CES.** El gDNA se fragmenta con transposones y mediante PCR se unen unos adaptadores que contienen el índice y una secuencia universal. Las librerías de distintas muestras se mezclan y se enriquecen en las regiones de interés con sondas unidas a biotina y bolas magnéticas con estreptavidina. Adaptada de [www.illumina.com](http://www.illumina.com)

La nomenclatura de las mutaciones se realizó sobre el transcrito de referencia siguiendo la normativa establecida por HGVS y para mayor exactitud se comprobó con el programa Mutalyzer 2.0. Las coordenadas genómicas siguen el ensamblaje GRCh37/hg19.

En resumen, en la **tabla 10** se comparan los distintos métodos de preparación de la muestra y las plataformas de análisis de datos empleadas en las aproximaciones de NGS utilizadas.

Tabla 10. Comparativa entre las aproximaciones de secuenciación masiva utilizadas.

	TES-CDG	TES-EMH	CES	WES
Número de genes a capturar	43	111	4.813	~ 20.000
Hibridación y captura	SureSelect (Agilent)	Haloplex (Agilent)	TruSight™ One (Illumina)	Human All Exon 50Mb o V4 (Agilent) o TruSeq (Illumina)
Fragmentación del DNA	Sonicación	Enzimas de restricción	Transposones	Sonicación
Clustering	PCR en emulsión	PCR en puente	PCR en puente	PCR en puente
Tamaño de la librería	0,2Mb	0,5Mb	12Mb	50-60Mb
Secuenciador	SOLiD4	MiSeq	MiSeq o NextSeq500	HiSeq2000
Método de secuenciación	Ligación	Síntesis	Síntesis	Síntesis
Longitud de las lecturas	2x50pb	2x150pb	2x250pb	2x75pb o 2x100pb
Plataforma de análisis de datos	DNA Nexus y BioMol	DNA Nexus	DNA Nexus, Variant Studio	BioMol, Genegrid y Variant Studio

## 2.4 Soporte informático y análisis *in-silico*

La significatividad estadística de las diferencias entre los grupos analizados se obtuvo mediante una prueba *T de Student* de dos colas, realizada mediante el programa *SPSS Statistics 2.1* (IBM). Cada valor representa la media y desviación estándar de al menos dos o tres conjuntos de datos. Las diferencias observadas se consideraron significativas según los p valores \* <0,5; \*\* <0,01; \*\*\* <0,001.

Además, a lo largo del trabajo se han empleado distintos soportes informáticos cuyo uso se resume en la **Tabla 7**.



# RESULTADOS

<b>1</b>	<b>SELECCIÓN Y ESTRATIFICACIÓN BIOQUÍMICA DE PACIENTES</b>	<b>61</b>
1.1	Selección de pacientes con sospecha de CDG	61
1.2	Selección de pacientes con sospecha de glucogenosis	63
<b>2</b>	<b>EVALUACIÓN DE DISTINTAS APROXIMACIONES DE SECUENCIACIÓN MASIVA</b>	<b>64</b>
2.1	Análisis de variantes detectadas por secuenciación masiva	64
2.1.1	Análisis de variantes en genes candidatos	64
2.1.2	Búsqueda de nuevos genes no asociados a patología	66
2.2	Paneles de captura personalizados (TES)	67
2.2.1	Panel personalizado de genes CDG	67
2.2.2	Panel de captura de genes de glucogenosis	67
2.3	Exoma completo (WES)	68
2.4	Exoma clínico (CES)	69
2.5	Comparativa entre las distintas aproximaciones de NGS empleadas	69
2.6	Detección de grandes deleciones	73
2.7	Análisis de lecturas de genes con pseudogenes	74
<b>3</b>	<b>APLICACIÓN DE LA NGS AL DIAGNÓSTICO DE PACIENTES</b>	<b>76</b>
3.1	Pacientes con defectos cogénitos de glicosilación	76
3.1.1	Identificación de un nuevo gen CDG	76
3.1.2	Espectro genético y mutacional de pacientes CDG	78
3.2	Pacientes con glucogenosis. Espectro genético y mutacional	84
3.3	Hallazgos inesperados	88

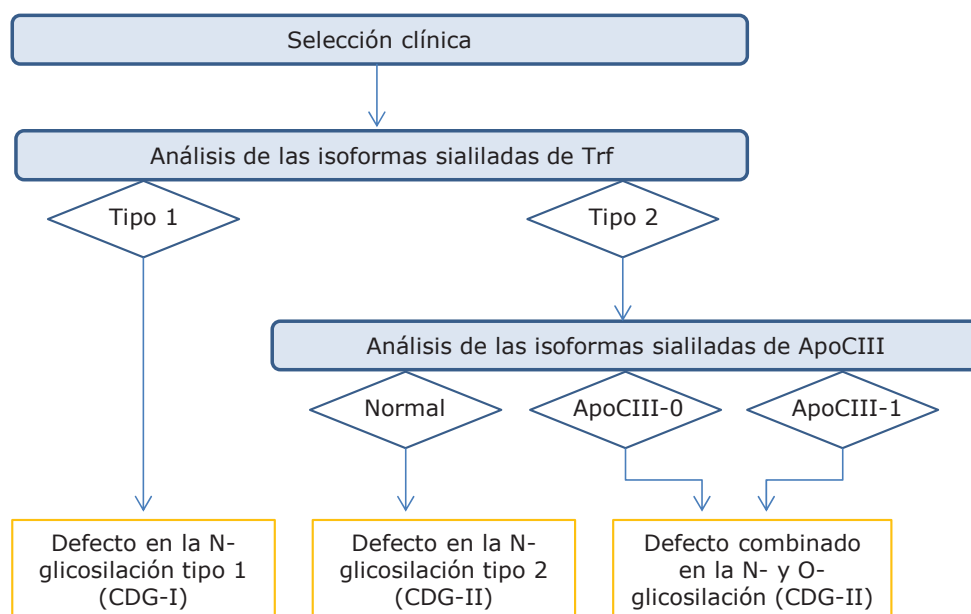
## 1 SELECCIÓN Y ESTRATIFICACIÓN BIOQUÍMICA DE PACIENTES

Ante una sospecha de una enfermedad metabólica hereditaria es necesario realizar el estudio de biomarcadores en fluidos biológicos del paciente. El perfil bioquímico obtenido puede proporcionar un indicio de la ruta metabólica afectada. La selección y estratificación de pacientes se realiza combinando la información clínica con los datos bioquímicos obtenidos.

### 1.1 Selección de pacientes con sospecha de CDG

La selección y clasificación de pacientes se realizó utilizando el siguiente flujo de trabajo diagnóstico (**Figura 10**). En primer lugar se analizó el patrón de isoformas de la proteína N-glicosilada transferrina por HPLC o isoelectroenfoque (IEF) como método de cribado de CDG a todos los pacientes con sospecha de enfermedad genética que presentaban una clínica multisistémica con afectación neurológica de origen inexplicado, referidos al Centro de Diagnóstico de Enfermedades Moleculares (CEDEM). En base al perfil anómalo de las isoformas sialiladas de la Trf se realizó una clasificación de los paciente en dos grupos: CDG-I y CDG-II tal y como se describe en el apartado 2.1.1 de materiales y métodos. De esta manera se seleccionaron 39 pacientes con sospecha de CDG, 17 de ellos se clasificaron como CDG-I y 21 como CDG-II. Además, se incluyó en el estudio otro paciente más que presentó un perfil de Trf normal pero que manifestaba una sintomatología clínica sugestiva de CDG (P44). En todos ellos se descartó la presencia de polimorfismos en la cadena polipeptídica de la transferrina como causa de ese patrón aberrante.

En los pacientes clasificados como CDG-II además se analizó el perfil de isoformas de la O-glicoproteína apolipoproteína CIII (ApoCIII). Aquellos que presentaron un patrón anómalo (apartado 2.1.3 de materiales y métodos) se subclasificaron en CDG-II con sospecha de un defecto combinado en la N- y O-glicosilación de proteínas. De todos los pacientes CDG-II analizados, dos presentaron un perfil anómalo; uno con un patrón tipo ApoCIII-0 y otro con un patrón ApoCIII-1.



**Fig 10.** Flujo de trabajo empleado para la clasificación de pacientes CDG. Se seleccionaron aquellos pacientes que presentaron una clínica multisistémica con afectación neurológica y se analizó el perfil de las isoformas de la N-glicoproteína transferrina, clasificándolos en CDG-I o CDG-II. En los CDG-II además se analizó el patrón de isoformas de la O-glicoproteína apolipoproteína CIII. Aquellos que también presentaron anomalías fueron subclasificados como “defectos combinados en la N- y O-glicosilación de proteínas”.

Además del análisis de los biomarcadores de glicosilación utilizados de manera estándar, en algunos CDG-II se analizaron las estructuras de los glicanos unidos a proteína con el fin de poder ayudar en el filtrado durante el estudio genético posterior. Se analizó la estructura de los N-glicanos unidos a proteínas plasmáticas por MALDI-TOF/MS tal y como se describe en el apartado 2.1.4 de materiales y métodos. El resultado de este estudio puede sugerir la enzima deficiente en la ruta de procesamiento del glicano N-unido. En total, se analizaron 9 pacientes con sospecha de CDG-II, obteniéndose valores elevados de estructuras anómalas de glicanos en 5 de ellos (**Tabla 11**).

**Tabla 11. Datos clínicos y porcentaje de glicanos anómalos de proteínas plasmáticas de pacientes CDG-II.** En granate se resaltan los parámetros elevados por encima de lo normal. VN, valores normales, indicado como la media  $\pm$  desviación estándar.

		P1	P2	P3	P4	P5
Sexo		F	F	F	F	F
Clínica		Retraso psicomental, hipotonía, encefalopatía, convulsiones, dismorfias	Retraso psicomotor, hipotonía, estrabismo, retraso mental, neuropatía periférica, hipotiroidismo	Encefalopatía, hepatopatía	Retraso psicomotor, hipotonía, convulsiones, síndrome de West	Hipotonía, dismorfias
% de glicanos (VN, n=8)	Truncados (42,6 $\pm$ 11,8)	76,9	70,0	68,0	64,8	68,1
	Tres antenas (11,2 $\pm$ 2,4)	17,7	16,9	18,3	20,9	12,2
	Gal terminal (31,6 $\pm$ 6,7)	46,6	42,5	46,0	37,5	44,0
	NAcGlc terminal (17,2 $\pm$ 8,8)	34,8	35,0	28,0	30,8	31,0
	Fucosa (28,2 $\pm$ 7,1)	32,8	39,3	42,8	28,0	33,1
	Baja sialización (29,1 $\pm$ 7,0)	69,7	55,8	45,7	44,21	50,8

F, femenino; Gal, galactosa; NAcGal, N-acetilgalactosamina.

## 1.2 Selección de pacientes con sospecha de glucogenosis

Los pacientes con sospecha de glucogenosis fueron seleccionados bien por indicación del médico solicitante o bien por presentar una clínica sugestiva de glucogenosis con afectación muscular y/o hepática y/o cardíaca.

Algunos pacientes además habían presentado evidencias de acúmulo de glucógeno a nivel muscular mediante histoquímica realizadas en el hospital de referencia. En total, se seleccionaron 47 pacientes con sospecha de glucogenosis.

## 2 EVALUACIÓN DE DISTINTAS APROXIMACIONES DE SECUENCIACIÓN MASIVA

---

En este apartado se describe la evaluación de tres aproximaciones de secuenciación masiva de DNA, un panel de genes personalizado (TES), el exoma celular completo (WES) y un panel amplio de genes asociados a enfermedad (CES), aplicados al diagnóstico de pacientes con sospecha de CDG y de glucogenosis. Además, se describe el flujo de trabajo establecido para analizar las variantes detectadas.

### 2.1 Análisis de variantes detectadas por secuenciación masiva

#### 2.1.1 Análisis de variantes en genes candidatos

Con el fin de filtrar y priorizar la búsqueda de variantes candidatas se estableció un protocolo a seguir en todas las aproximaciones de NGS empleadas (**Figura 11**).

Se establecieron parámetros de calidad para determinar las variantes (SNV) suficientemente robustas para continuar con el análisis, descartando aquellas con menos de diez lecturas ( $DC < 10x$ ) y menos de un 15% de variación.

Se filtraron los SNV para que tuviesen una frecuencia alélica detectada en la población control menor de 1% según las bases de datos 1000g, EVS y ExAc. Las variantes localizadas en regiones intrónicas internas y UTR se excluyeron del análisis. Además, cuando la plataforma de análisis de variantes lo permitía (sólo *Variant Studio*), se realizó un cruzado de datos con una muestra analizada en el mismo ensayo con el fin de eliminar variantes que se repiten en las dos muestras y que serán variantes polimórficas o debidas a errores recurrentes del proceso de secuenciación masiva.

Para los análisis por CES o WES se procedió a una priorización de las variantes analizando aquellas localizadas en genes descritos asociadas a la patología a estudio. Si tras esta primera priorización no se encontraron cambios candidatos, se procedió a una segunda priorización en genes asociados a otras patologías o no asociados a enfermedad hasta la fecha. Para los CDG se priorizó una lista de 150 genes candidatos implicados no sólo en la ruta de N-glicosilación, sino también en otras rutas del metabolismo de carbohidratos así como genes codificantes para proteínas implicadas en tráfico vesicular (**Anexo 5**). Para la glucogenosis se priorizaron los genes descritos asociados a la enfermedad (**Anexo 4**).

Se priorizaron las variantes de tipo LoF (*Loss of function*), es decir, inserciones y/o deleciones o duplicaciones que producen una pérdida en la pauta de lectura (*frameshift*), cambios que generan un codón de parada (*nonsense*), pérdida de los sitios naturales de *splicing* 3' y 5' *splice site* o pérdida del codón de inicio o de parada. Además, se comprobó si las variantes estaban anotadas en las bases de datos de mutaciones en humanos (*HGMD® professional*). A continuación se valoró la posible patogenicidad de las variantes a partir de la consecuencia predecible utilizando los programas de predicción *in-silico* detallados a continuación.

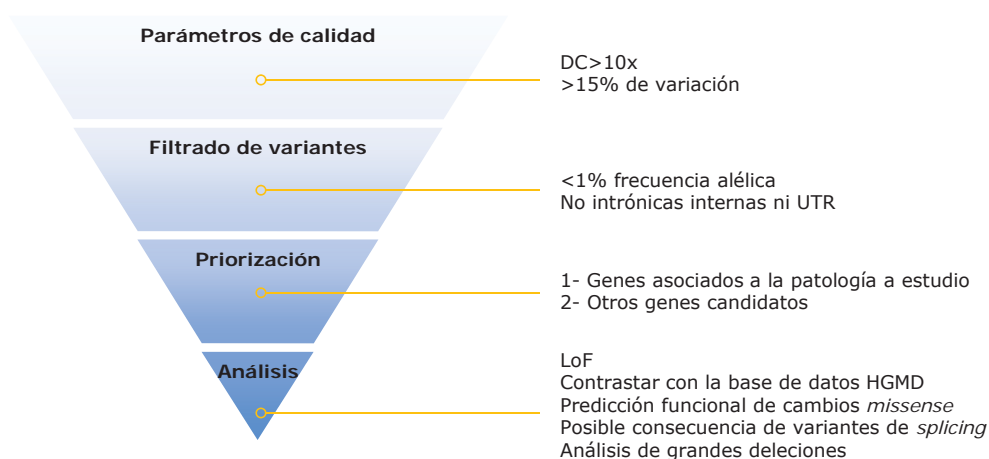
La predicción funcional de los cambios nucleotídicos que pudieran originar un cambio aminoacídico se realizó con SIFT, *Polyphen2* y *Mutation Taster*, además de otros predictores incluidos en las plataformas de análisis de datos de NGS. Para valorar si estos estaban localizados en el péptido señal de la proteína y pudieran afectar a la correcta localización subcelular de la misma se usó SignalIP. También se analizó la conservación entre especies del aminoácido afectado realizando un Clustal Omega. Asimismo, utilizando la información contenida en las bases de datos PDB y UniProt se estudió la interacción del aminoácido afectado con otros aminoácidos de la proteína en su estructura cuaternaria o con cofactores, iones o el propio sustrato.

Para el análisis de las variantes que podrían afectar a los sitios naturales de *splicing* (5' y 3' *splice site*) se usaron los programas BDGP, HSF y AST. Para variantes exónicas que podrían afectar a dominios de unión de proteínas implicadas en *splicing* se usó *ESE finder* y para variantes intrónicas que podrían activar la inserción de un pseudoexón se utilizó *Cryp Skip*.

Cuando no se detectó ningún posible cambio patogénico se realizó una búsqueda más amplia incluyendo variantes con una DC>5% y aquellas variantes detectadas en las regiones 3'UTR de los genes candidatos. Para estos últimos, se analizó si afectaban a la diana de algún miRNA implicado en la estabilidad del mRNA mediante los programas de predicción *miRror Suite* y PITA.

También, se realizó un estudio de grandes deleciones a partir de los datos de profundidad de cobertura tal y como se describe en el apartado 2.6 de resultados. Además, se analizó la cobertura horizontal de todos los genes descritos asociados a la patología a estudio y se secuenciaron por Sanger las regiones poco cubiertas. Todas las variantes candidatas se confirmaron por secuenciación por Sanger en el caso índice y, siempre que fue posible, se analizó la segregación mendeliana en la familia.

Todos los programas y bases de datos utilizados en el análisis de datos se encuentran recopilados en la **Tabla 7**.



**Fig 11.** Protocolo para el análisis de variantes de secuenciación masiva. En primer lugar se establecieron unos parámetros de calidad de las variantes para proceder a un filtrado según la localización de las SNV en el genoma y según su frecuencia alélica poblacional. Posteriormente se hizo un priorizado de SNV localizados en los genes de interés para finalmente realizar un análisis según su posible patogenicidad a partir de la consecuencia predecible.

### 2.1.2 Búsqueda de nuevos genes no asociados a patología

Cuando el análisis de datos de WES no aportó ningún resultado, se procedió a un estudio más amplio en busca de nuevos genes candidatos no asociados a la glicosilación de proteínas hasta la fecha. Para ello se amplió la búsqueda a todos los genes manteniendo los parámetros de calidad de las variantes establecidos anteriormente (DC>10x, variación >15%) pero se siguió un filtrado posterior más exigente.

Se seleccionaron las variantes con una frecuencia alélica detectada en población sana <0,05% y se excluyeron aquellas localizadas en las regiones UTR, intrónicas más alejadas de  $\pm 6$  nucleótidos respecto del exón y sinónimas.

Se priorizaron las variantes de tipo *frameshift* y *nonsense* y aquellas *missense* predichas como patogénicas por al menos seis de los siete programas de predicción *in-silico* utilizados por la plataforma de trabajo (SIFT *protein*, SIFT MSA, PhyloP, Polyphen, LRT, *Mutation taster* y GERP++). Además, se supuso un modo de herencia autosómico recesivo. Finalmente, en base a la función del producto génico según las bases de datos de Uniprot, GeneCards y *Gene Ontology Function*, la ruta celular en la que participan según *Gene Ontology Pathways*, *Kegg pathway* y *BioCarta Pathways*, su localización subcelular según las bases de datos BioGPS, GeneHub GEPIS y *Human Protein Atlas* y otras proteínas con las que interacciona según *String*, se seleccionaron aquellos genes que pudiesen estar relacionados con la ruta de glicosilación. De esta manera, se generó un listado por paciente de 10-15 genes candidatos con cambios potencialmente patogénicos.

## 2.2 Paneles de captura personalizados (TES)

### 2.2.1 Panel personalizado de genes CDG

Para el análisis mediante un panel personalizado de genes CDG (TES-CDG) se seleccionaron 16 pacientes con sospecha de CDG, 7 de ellos clasificados como CDG-I y 9 como CDG-II. Entre los CDG-II incluidos en este estudio se encontraban los 5 pacientes que habían presentado acumulación de estructuras anómalas de glicanos y uno de los pacientes con sospecha de un defecto combinado en la N- y O-glicosilación de proteínas.

Se diseñó un panel de 43 genes implicados en la N-glicosilación de proteínas descritos hasta entonces (**Anexo 3**). Se capturaron las regiones codificantes, 25pb de las regiones intrónicas flanqueantes y las UTR de los genes de interés. Se registró un promedio de 70-75 SNV por paciente que se filtraron para su análisis. La profundidad de cobertura media por gen fue de un 50x y por exón fue siempre mayor de 10x.

Se detectaron cambios probablemente patogénicos en los genes *RFT1* y *DPAGT1* en dos de los pacientes CDG-I analizados y una única variante posiblemente patogénica en heterocigosis en el gen *COG7* en uno de los pacientes CDG-II. La tasa de éxito del panel fue de un 19% (3/16).

### 2.2.2 Panel de captura de genes de glucogenosis

Para el estudio de pacientes con sospecha de glucogenosis se seleccionaron un total de 16 pacientes y se utilizó un panel personalizado (TES-EMH) diseñado para capturar las regiones codificantes y 25pb de las regiones intrónicas flanqueantes junto con las UTRs de 111 genes relacionados con enfermedades metabólicas hereditarias, entre ellos, 22 genes de glucogenosis (**Anexo 4**).

La cobertura mínima alcanzada fue de 20x para el 95% de las bases de interés, excepto para *PHKA2* y *GBE1* que fue del 85%. Todas las regiones poco cubiertas (<10x) pertenecían a secuencias intrónicas. Se detectó un promedio de 1.200 SNV por paciente.

Se detectaron tres pacientes que portaban variantes bialélicas previsiblemente patogénicas en los genes de glucogenosis *PYGL*, *GAA* y *AGL*. Además, se detectó otro paciente heterocigoto para una mutación previamente descrita en el gen *PHKB*.

Puesto que el panel incluía 111 genes relacionados con enfermedades metabólicas hereditarias, también se detectó un paciente con una mutación en homocigosis en el gen *ALDOB*. En resumen, este panel permitió la detección de cambios potencialmente patogénicos en 5 pacientes, originando una tasa diagnóstica del 31% (5/16).



## 2.3 Exoma completo (WES)

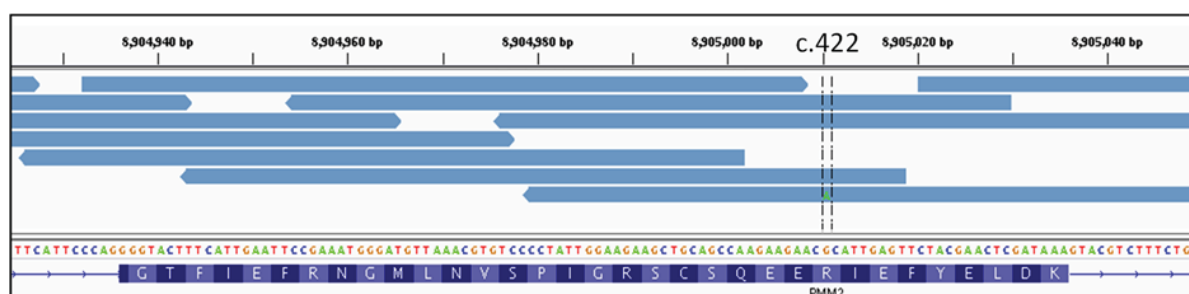
El estudio del exoma completo incluye el análisis de los exones de los genes que codifican para proteínas y 50pb de las regiones intrónicas flanqueantes. El análisis del exoma completo se realizó en 16 pacientes con sospecha de CDG, 7 de ellos clasificados como CDG-I y 9 como CDG-II. Cinco de los pacientes habían sido analizados por TES-CDG sin obtener un diagnóstico genético.

La cobertura media de los genes implicados en la ruta de la N-glicosilación de proteínas descritos en el **Anexo 3** fue de 70x, siendo la mínima detectada de 33x en los genes *COG8* y *SLC35C1*. Se detectó un promedio de 45.000 SNV por paciente. Mediante el análisis por WES se detectaron cambios previsiblemente patogénicos en genes CDG previamente descritos en 6 pacientes (*DPAGT1*, *DPM1*, *DOLK*, *GALT*, *PMM2* y *SSR4*).

El análisis exhaustivo de los SNV detectados en busca de nuevos genes relacionados con la ruta de glicosilación nos permitió la identificación, en colaboración con el grupo del Dr. Lefeber (Nimega, Holanda), de una mutación en homocigosis en un nuevo gen no relacionado con ninguna patología humana hasta la fecha, *CCDC115*. Por tanto, la tasa diagnóstica de WES fue de un 44% (7/16).

En el caso del paciente PMM2-CDG (P20), se detectó una única mutación en heterocigosis previamente descrita (c.95T>G) en el gen *PMM2*. Dado que la clínica y la bioquímica del paciente eran consistentes con un defecto en este gen, se secuenció por Sanger convencional detectándose una segunda mutación (c.422G>A). El análisis de la profundidad de cobertura a partir de los archivos BAM mostró una cobertura igual o inferior a 6x en todos los nucleótidos de ese exón. En la posición c.422 se identificaron cuatro lecturas, una de ellas con la mutación (**Figura 12**).

Basándonos en las características clínicas y bioquímicas del paciente deficiente en *GALT*, reevaluamos los datos clínicos de los pacientes CDG no caracterizados y mediante secuenciación por Sanger convencional identificamos dos más deficientes en este gen.



**Fig 12.** Profundidad de cobertura del exón 5 del gen *PMM2* del P20. En la parte superior se indican las coordenadas genómicas en hg19, en el centro las lecturas alineadas y en la parte inferior la secuencia nucleotídica y aminoacídica. La posición de la variante no detectada durante el análisis de los datos de WES (c.422) se indica con dos líneas discontinuas verticales.

## 2.4 Exoma clínico (CES)

El panel de exoma clínico comercial empleado captura los exones y 50pb flanqueantes de 4.813 genes asociados a enfermedad anotados en la base de datos de OMIM. La cobertura mínima fue de un 20x en el 97% de las bases de interés de los genes CDG utilizados en el filtrado (**Anexo 5**), y en el 99% de genes de glucogenosis (**Anexo 4**). La cobertura media fue de 74x en ambos casos y se detectó un promedio de 8.200 SNV por paciente.

El análisis del exoma clínico se realizó en 27 pacientes con sospecha de CDG (7 de ellos previamente analizados mediante el panel TES-CDG y 8 por WES, todos sin diagnóstico genético), y se detectaron mutaciones causantes de enfermedad en 8 pacientes en los genes *ALG1*, *ALG6*, *B4GALT1*, *GFPT1*, *PMM2* o *SRD5A3*. Además, reanalizando los datos buscando otros genes relacionados con la patología del paciente se detectaron dos mutaciones causantes de enfermedad en el gen *FOLR1*. Así pues, la tasa diagnóstica de CES para el conjunto de pacientes de CDG fue de un 33% (9/27).

También se analizaron mediante CES 43 pacientes con sospecha de glucogenosis (12 de ellos previamente analizados por TES-EMH), detectándose variantes posiblemente patogénicas en tres genes relacionados con glucogenosis (*AGL*, *PHKA2* y *SLC37A4*) en 14 pacientes. Asimismo, se detectaron mutaciones en otros cuatro genes no relacionados con glucogenosis (*LIPA*, *CPT2*, *ANO5* y *NKX2-5*), cuyo defecto produce características fenotípicas solapantes. La tasa diagnóstica de CES para el conjunto de pacientes de glucogenosis fue de un 42% (18/43).

## 2.5 Comparativa entre las distintas aproximaciones de NGS empleadas

En resumen, mediante las distintas aproximaciones de NGS empleadas, se han caracterizado genéticamente un 49% (19/39) de los pacientes con sospecha de CDG, y otro 49% (23/47) de las sospechas de glucogenosis (**Figura 13**).

La profundidad de cobertura obtenida por todas las aproximaciones de NGS empleadas se comparó en los genes de CDG o glucogenosis comunes a todas ellas (**Tabla 12**). La profundidad de cobertura de TES-CDG fue siempre la más baja, excepto para los genes *COG8*, *DDOST* y *SLC35C1*, que fue ligeramente superior a la alcanzada por WES. Por lo general, la más alta se registró en CES, aunque en algunos genes fue superado ligeramente por la alcanzada por WES.

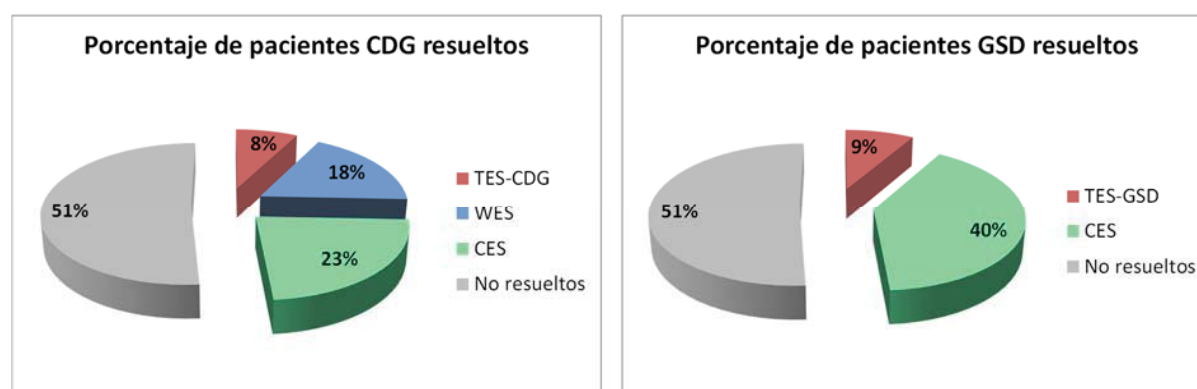
Respecto a los genes de glucogenosis, la profundidad de cobertura fue siempre muy superior en TES-EMH aunque los datos de la desviación estándar resaltan la falta de homogeneidad entre las muestras (**Tabla 13**).

**Tabla 12. Comparativa de la profundidad de cobertura media de los genes CDG analizados mediante TES-CDG, WES y CES.** Los datos indican el valor medio  $\pm$  su desviación estándar. n es el número de pacientes utilizados para el cálculo de la profundidad de cobertura.

Gen	Transcrito	MIM	Profundidad de cobertura		
			TES-CDG (n=16)	WES (n=6)	CES (n=10)
<i>ALG1</i>	NM_019109	605907	43 $\pm$ 2	44 $\pm$ 7	69 $\pm$ 13
<i>ALG2</i>	NM_033087	607906	45 $\pm$ 2	67 $\pm$ 7	92 $\pm$ 20
<i>ALG3</i>	NM_005787	608750	48 $\pm$ 3	44 $\pm$ 6	85 $\pm$ 16
<i>ALG6</i>	NM_013339	604566	50 $\pm$ 3	95 $\pm$ 13	80 $\pm$ 19
<i>ALG8</i>	NM_024079	608103	50 $\pm$ 3	72 $\pm$ 9	58 $\pm$ 14
<i>ALG9</i>	NM_024740	606941	47 $\pm$ 2	81 $\pm$ 11	66 $\pm$ 14
<i>ALG11</i>	NM_001004127	613666	28 $\pm$ 1	120 $\pm$ 13	114 $\pm$ 26
<i>ALG12</i>	NM_024105	607144	51 $\pm$ 3	38 $\pm$ 5	72 $\pm$ 14
<i>ATP6V0A2</i>	NM_012463	611716	51 $\pm$ 2	82 $\pm$ 11	68 $\pm$ 13
<i>B4GALT1</i>	NM_001497	137060	42 $\pm$ 2	38 $\pm$ 6	57 $\pm$ 11
<i>COG1</i>	NM_018714	606973	53 $\pm$ 2	76 $\pm$ 8	67 $\pm$ 13
<i>COG4</i>	NM_015386	606976	55 $\pm$ 3	60 $\pm$ 8	73 $\pm$ 15
<i>COG5</i>	NM_006348	606821	45 $\pm$ 3	73 $\pm$ 9	70 $\pm$ 16
<i>COG6</i>	NM_020751	606977	45 $\pm$ 3	60 $\pm$ 9	71 $\pm$ 16
<i>COG7</i>	NM_153603	606978	55 $\pm$ 3	58 $\pm$ 6	67 $\pm$ 12
<i>COG8</i>	NM_032382	606979	45 $\pm$ 3	34 $\pm$ 4	64 $\pm$ 13
<i>DDOST</i>	NM_005216	602202	54 $\pm$ 3	42 $\pm$ 7	76 $\pm$ 14
<i>DHDDS</i>	NM_024887	608172	55 $\pm$ 3	84 $\pm$ 10	55 $\pm$ 10
<i>DOLK</i>	NM_014908	610746	52 $\pm$ 3	82 $\pm$ 12	111 $\pm$ 18
<i>DPAGT1</i>	NM_001382	191350	56 $\pm$ 3	96 $\pm$ 14	70 $\pm$ 13
<i>DPM1</i>	NM_003859	603503	46 $\pm$ 3	95 $\pm$ 15	85 $\pm$ 22
<i>DPM3</i>	NM_153741	605951	45 $\pm$ 3	40 $\pm$ 4	53 $\pm$ 11
<i>GNE</i>	NM_005476	603824	55 $\pm$ 2	92 $\pm$ 11	68 $\pm$ 12
<i>MPDU1</i>	NM_004870	604041	53 $\pm$ 3	67 $\pm$ 8	67 $\pm$ 13
<i>MPI</i>	NM_002435	154550	58 $\pm$ 3	63 $\pm$ 8	103 $\pm$ 18
<i>PMM2</i>	NM_000303	601785	52 $\pm$ 3	77 $\pm$ 8	62 $\pm$ 15
<i>RFT1</i>	NM_052859	611908	50 $\pm$ 3	64 $\pm$ 9	64 $\pm$ 13
<i>SLC35A1</i>	NM_006416	605634	48 $\pm$ 3	67 $\pm$ 10	54 $\pm$ 14
<i>SLC35C1</i>	NM_018389	605881	48 $\pm$ 3	34 $\pm$ 4	103 $\pm$ 24
<i>SRD5A3</i>	NM_024592	611715	45 $\pm$ 3	90 $\pm$ 12	63 $\pm$ 15
<i>TUSC3</i>	NM_006765	601385	43 $\pm$ 2	92 $\pm$ 12	85 $\pm$ 20

**Tabla 13. Comparativa de las coberturas medias de los genes GSD analizados mediante TES-EMH y CES.** Los datos indican el valor medio  $\pm$  su desviación estándar. n es el número de pacientes empleados para el cálculo de la profundidad de cobertura.

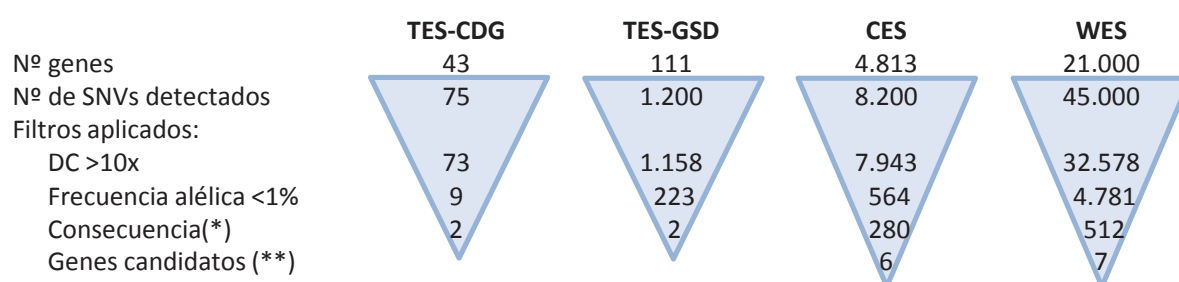
Gen	Transcrito	MIM	Profundidad de cobertura	
			TES-EMH (n=12)	CES (n=41)
<i>AGL</i>	NM_000642	610860	173 $\pm$ 89	75 $\pm$ 19
<i>ALDOA</i>	NM_000034	103850	455 $\pm$ 206	96 $\pm$ 16
<i>ENO3</i>	NM_053013	131370	582 $\pm$ 264	78 $\pm$ 13
<i>G6PC</i>	NM_000151	613742	439 $\pm$ 185	86 $\pm$ 17
<i>GAA</i>	NM_000152	606800	595 $\pm$ 283	90 $\pm$ 16
<i>GBE1</i>	NM_000158	607839	220 $\pm$ 114	80 $\pm$ 22
<i>GYG1</i>	NM_004130	603942	265 $\pm$ 124	73 $\pm$ 16
<i>GYS1</i>	NM_002103	138570	585 $\pm$ 266	94 $\pm$ 17
<i>GYS2</i>	NM_021957	138571	252 $\pm$ 131	87 $\pm$ 23
<i>LDHA</i>	NM_005566	150000	235 $\pm$ 104	90 $\pm$ 18
<i>PFKM</i>	NM_000289	610681	543 $\pm$ 230	76 $\pm$ 15
<i>PGAM2</i>	NM_000290	612931	525 $\pm$ 288	104 $\pm$ 21
<i>PGM1</i>	NM_002633	171900	570 $\pm$ 233	81 $\pm$ 18
<i>PHKA1</i>	NM_002637	311870	185 $\pm$ 93	87 $\pm$ 42
<i>PHKA2</i>	NM_000292	300798	369 $\pm$ 201	85 $\pm$ 42
<i>PHKB</i>	NM_000293	172490	269 $\pm$ 135	76 $\pm$ 21
<i>PHKG2</i>	NM_000294	172471	453 $\pm$ 195	60 $\pm$ 11
<i>PYGL</i>	NM_002863	613741	459 $\pm$ 187	68 $\pm$ 15
<i>PYGM</i>	NM_005609	608455	723 $\pm$ 345	100 $\pm$ 25
<i>SLC2A2</i>	NM_000340	138160	416 $\pm$ 171	81 $\pm$ 18
<i>SLC37A4</i>	NM_001164277	602671	921 $\pm$ 415	93 $\pm$ 17



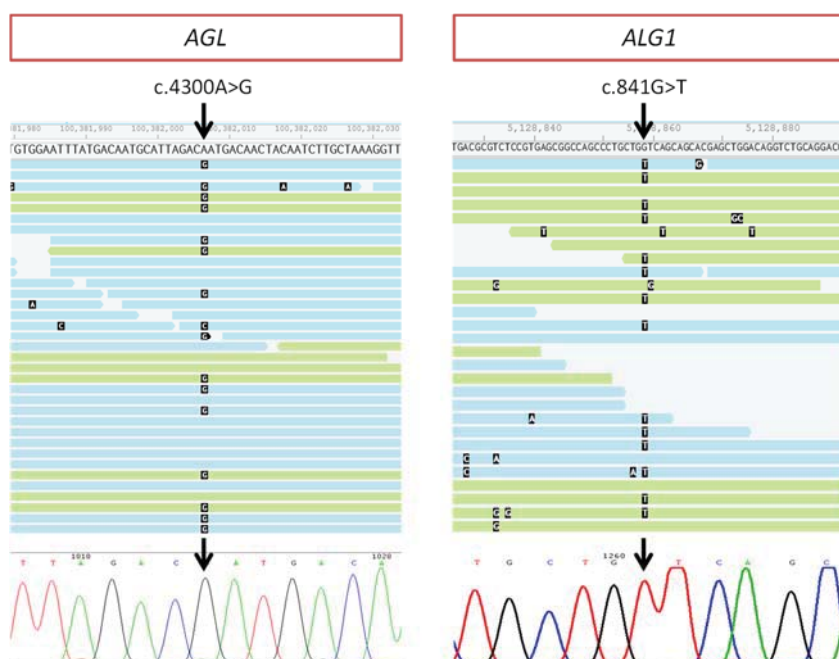
**Fig 13.** Porcentaje de pacientes resueltos por secuenciación masiva. Conjunto de pacientes con sospecha de CDG (izquierda) o de glucogenosis (derecha).

El número de variantes detectadas mediante cada aproximación es proporcional a la región capturada en el diseño. La aplicación de los parámetros de calidad y el filtrado de variantes descrito en el apartado 2.1 de resultados fue reduciendo el número de SNV (**Figura 14**). El porcentaje de SNV que se descartó en el filtrado por frecuencia alélica fue similar en todas las aproximaciones empleadas (93-81%). Sin embargo, el filtrado según la profundidad de cobertura (DC>10x) excluyó más SNV en WES (28% frente al 3% en las demás aproximaciones) y según la consecuencia descartó más SNV en TES-EMH (99% frente al 88-50%).

Cabe resaltar la identificación mediante CES de una variante en heterocigosis en *AGL* (44,4% de las lecturas mostraban el cambio) cuya comprobación por Sanger convencional mostró, sin embargo, la variante en homocigosis. El empleo de distintas parejas de oligonucleótidos por Sanger permitió finalmente detectar la heterocigosidad de la variante. Este efecto es conocido como *allele dropout*, referente a la falta de amplificación de uno de los alelos. Lo mismo ocurrió en el estudio familiar por Sanger de un caso de CDG, donde uno de los progenitores presentó la variante patogénica en homocigosis en *ALG1*. La secuenciación por CES mostró la heterocigosidad de la variante (41,6% de las lecturas portaban el cambio), apuntando de nuevo a un *allele dropout* en Sanger (**Figura 15**).



**Fig 14. Filtrado de variantes obtenidas por secuenciación masiva.** En la figura se ejemplifica el número de variantes obtenidas tras la aplicación de cada filtro. (\*) Deleciones e inserciones exónicas, *frameshift*, mutaciones que afectan al codón de iniciación o terminación, *nonsense*, *splicing*, *missense*. (\*\*) Priorización de genes candidatos aplicable a CES o WES.



**Fig 15.** *Allele dropout* en la secuenciación por Sanger convencional. La parte superior de la figura muestra el gen y el cambio nucleotídico. La zona central ilustra las lecturas alineadas, donde la variante se resalta con un cuadrado negro. La parte inferior muestra la secuencia por Sanger, donde el cambio se indica con una flecha.

## 2.6 Detección de grandes deleciones

Cuando se detectó un único cambio en heterocigosis o para cambios aparentemente homocigotos, se comparó la profundidad de cobertura de esa muestra con muestras analizadas en el mismo ensayo con el fin de detectar grandes deleciones. Este análisis sólo se puede realizar en las aproximaciones que tienen suficiente profundidad de cobertura y que además sea homogénea (TES o CES). Para ello, se dividió la profundidad de cobertura media normalizada por exón del sujeto a estudio entre la media de varias muestras. Si la razón era de 0,5 se sospechaba de una deleción en heterocigosis, si era cero en homocigosis, y uno, no había deleción. Para comprobar que la disminución en la profundidad de cobertura se debía a una pérdida de dosis se empleó una técnica específica para la detección de deleciones, el SNP-array.

Procediendo de esta manera, se analizaron 6 pacientes del conjunto de glucogenosis y 9 de CDG en los que se había detectado una mutación aparentemente en homocigosis. En uno de los pacientes de CDG (P6) se identificó una deleción en heterocigosis en el gen *ALG1* que comenzaba en el exón 4 hasta el final del gen (**Tabla 14**). La deleción se comprobó por array de SNP y abarca desde el exón 4 de *ALG1* hasta la región 3' *downstream* del gen *EEF2KMT*, gen que también se delecciona (**Figura 16**).

**Tabla 14. Análisis de la profundidad de cobertura para la detección de grandes deleciones.** En rojo se resaltan los exones delecionados.

ALG1 (NM_019109)				Profundidad de cobertura media normalizada				
Exón	Coordenadas (hg19)		Longitud	Ct 1	Ct 2	Ct 3	P6	Cociente
1	5.121.820	5.122.058	228	0,31	0,43	0,51	0,42	1,0
2	5.122.952	5.123.029	98	0,82	0,79	0,70	1,02	1,3
3	5.123.154	5.123.257	124	0,99	1,07	1,14	0,87	0,8
4	5.125.389	5.125.537	169	0,80	0,75	0,73	0,30	0,4
5	5.127.446	5.127.535	110	1,10	1,27	1,01	0,69	0,6
6	5.127.908	5.128.018	131	0,55	0,52	0,70	0,31	0,5
7	5.128.758	5.128.879	142	0,69	0,79	0,67	0,43	0,6
8	5.129.065	5.129.103	59	0,89	0,95	0,74	0,37	0,4
9	5.129.749	5.129.808	80	0,68	0,48	0,47	0,31	0,6
10	5.130.947	5.131.057	131	0,52	0,73	0,49	0,29	0,5
11	5.132.560	5.132.674	135	0,65	0,75	0,70	0,41	0,6
12	5.133.683	5.133.758	96	1,07	1,14	1,11	0,51	0,5
13	5.134.751	5.135.395	152	1,49	1,64	1,44	0,97	0,6



**Fig 16. Región delecionada en el P6.** En la parte superior se indican las coordenadas genómicas en hg19. En la parte inferior se esquematizan los genes localizados en esa región, con los cuadrados representando los exones y las líneas finas los intrones. Las líneas verticales verde y azul de la figura indican los primeros SNP detectados en homocigosis en el SNParray (rs4786589 y rs4786614).

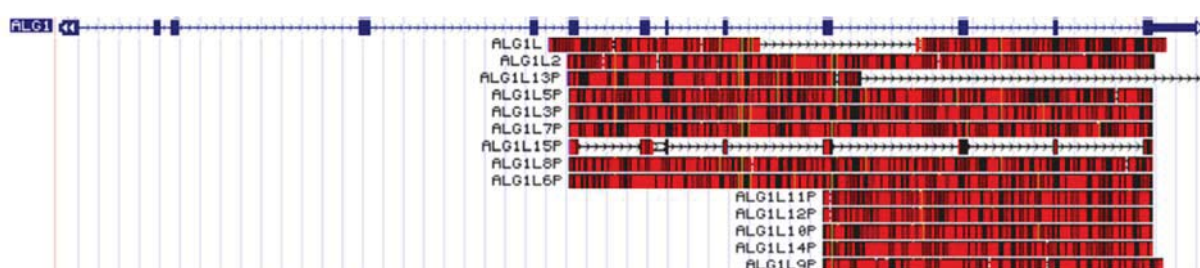
## 2.7 Análisis de lecturas de genes con pseudogenes

Uno de los mayores obstáculos de la NGS es el alineamiento de lecturas provenientes de genes con alta homología de secuencia a otras regiones del genoma, ya que los programas bioinformáticos empleados en el alineamiento encuentran dificultades para discernir con precisión las secuencias de los genes funcionales de las de los pseudogenes. Además, estas lecturas pueden ser descartadas del análisis ya que no alinean en una región única del genoma.

*ALG1*, uno de los genes CDG, posee alta homología de secuencia a dos genes codificantes de proteína (*ALG1L* y *ALG1L2*) y 12 pseudogenes (*ALG1L3P*, *ALG1L5-ALG1L15P*). Todos ellos son homólogos en su secuencia en la región que se extiende desde el exón 6 hasta el final de *ALG1* (Figura 17).

Se han detectado dos pacientes con cambios en *ALG1* y en ambos casos ha sido mediante CES. Se valoró la posibilidad de que las SNV identificadas (c.841G>T, c.854G>A y c.1228T>C) perteneciesen a lecturas de los pseudogenes mal alineadas en *ALG1* y que realmente se tratarasen de variantes naturales de los pseudogenes. Para ello se diseñaron oligonucleótidos específicos de *ALG1* en regiones donde no hubiese tanta homología de secuencia con los pseudogenes y se hizo una PCR a temperaturas de anillamiento altas, comprobándose de esta manera la presencia de las mutaciones en *ALG1*.

Curiosamente, la mutación c.841 G>T (p.Val281Phe) identificada mediante CES, no fue detectada por TES-CDG ni WES a pesar de que el gen estaba incluido en ambos diseños. Un análisis de los archivos BAM de TES y WES permitió observar una disminución de la DC en aquellos exones homólogos a otras regiones del genoma, probablemente debido a una mala calidad de sus lecturas que fueron descartadas al alinear en diferentes regiones del genoma.



**Fig 17.** Alineamiento del gen *ALG1* con genes y pseudogenes homólogos a su secuencia. En azul en la parte superior se indica el gen *ALG1* con sus exones representados como cuadrados y los intrones como líneas. En la parte inferior se dibujan los genes y pseudogenes con homología de secuencia a *ALG1*.



### 3 APLICACIÓN DE LA NGS AL DIAGNÓSTICO DE PACIENTES

#### 3.1 Pacientes con defectos congénitos de glicosilación

##### 3.1.1 Identificación de un nuevo gen CDG

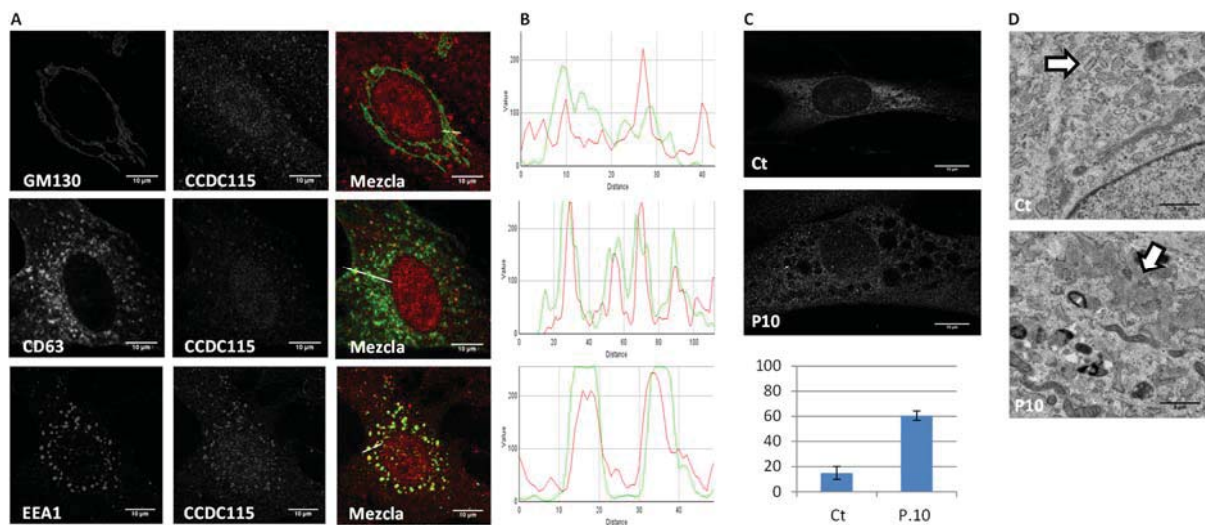
Los datos obtenidos por WES de los pacientes sin mutaciones en genes candidatos se filtraron con el fin de identificar posibles genes candidatos no vinculados a ninguna patología hasta la fecha. Estos datos fueron compartidos junto con otros grupos que trabajan en CDG en Europa encontrando un mismo gen candidato en uno de nuestros pacientes y siete del grupo del Dr. Lefeber (Nimega, Holanda).

El P10 presentaba problemas de aprendizaje, distribución anómala de la grasa y hepatopatía. Las transaminasas (AST y ALT), el colesterol y la fosfatasa alcalina estaban elevados en plasma mientras que la ceruloplasmina estaba disminuida. La biopsia hepática mostró niveles elevados de cobre. El estudio de glicoproteínas plasmáticas mostró un patrón tipo 2 en el Trf-IEF y un patrón ApoCIII-1 en el ApoCIII-IEF, lo que apunta a un CDG-II debido a un defecto combinado en la N- y O-glicosilación de proteínas (Jansen et al., 2016b). La evaluación del transporte retrógrado en los fibroblastos derivados del P10 mostró un resultado comparable a los controles.

El P10 presentó el cambio p.Leu31Ser en homocigosis en el gen *CCDC115* (NM\_032357.3). La posición Leu31 está conservada entre especies incluso alejadas filogenéticamente y el cambio por Serina es predicho como patogénico por todos los programas de predicción *in-silico* utilizados. Además, se comprobó la segregación mendeliana del cambio en la familia. No obstante, el nivel de proteína inmunorreactiva analizado por *Western Blot* mostró un resultado comparable al control utilizado.

La proteína CCDC115 muestra homología de secuencia con Vma22p de levadura, un factor de ensamblaje de la ATPasa-V del RE. Con el fin de determinar la localización subcelular de la proteína CCDC115 se realizaron ensayos de inmunofluorescencia en fibroblastos control con marcadores de CCDC115 y de distintos orgánulos celulares, detectándose una colocalización parcial con el marcador de endosomas tempranos EEA1 (**Figura 18 A-B**). Si bien, no todo el marcaje de CCDC115 colocaliza con EEA1, lo que parece indicar que la proteína a estudio también se ubica en otras vesículas. Estudios complementarios realizados en colaboración con el grupo del Dr. Lefeber mostraron una colocalización con el marcador del compartimento intermedio RE-AG (ERGIC53) y con las vesículas del transporte retrógrado AG-RE COPI (Beta-COP). Además, se observó una colocalización parcial con las vesículas de transporte anterógrado RE-AG COPII (Sec31) (Jansen et

al., 2016b). Estos resultados parecen indicar una localización de CCDC115 en vesículas. Además, mediante inmunofluorescencia (**Figura 18 C**) o microscopía electrónica (**Figura 18 D**) se observó una estructura dilatada y vacuolizada del RE en un 60% de las células del P10, comparada con el 15% de los fibroblastos control.

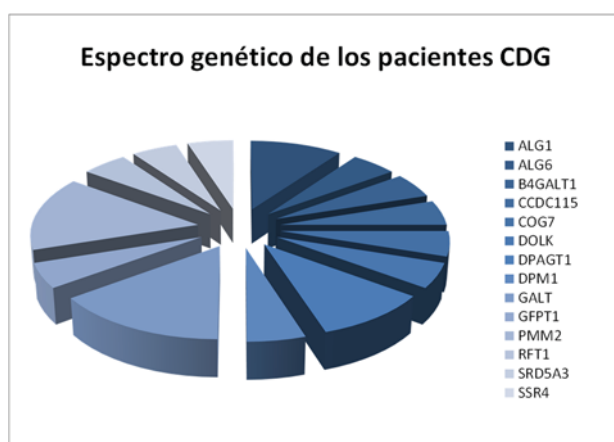


**Fig 18.** Caracterización molecular de CCDC115. **(A)** Estudios de colocalización de CCDC115 (rojo) con marcadores de AG (GM130), lisosomas (CD63) y endosomas tempranos (EEA1) (verde). Escala 10µm. **(B)** Perfiles de la intensidad de fluorescencia de cada fluoróforo de la figura "A" a lo largo de la línea blanca dibujada en la mezcla de los dos canales. **(C)** Estructura del RE de unos fibroblastos control (Ct) y de los fibroblastos derivados del paciente. Escala 10µm. El gráfico muestra el porcentaje de células con RE dilatado. **(D)** Análisis ultraestructural del RE de unos fibroblastos control (Ct) y de los fibroblastos derivados del paciente. Escala 5µm.

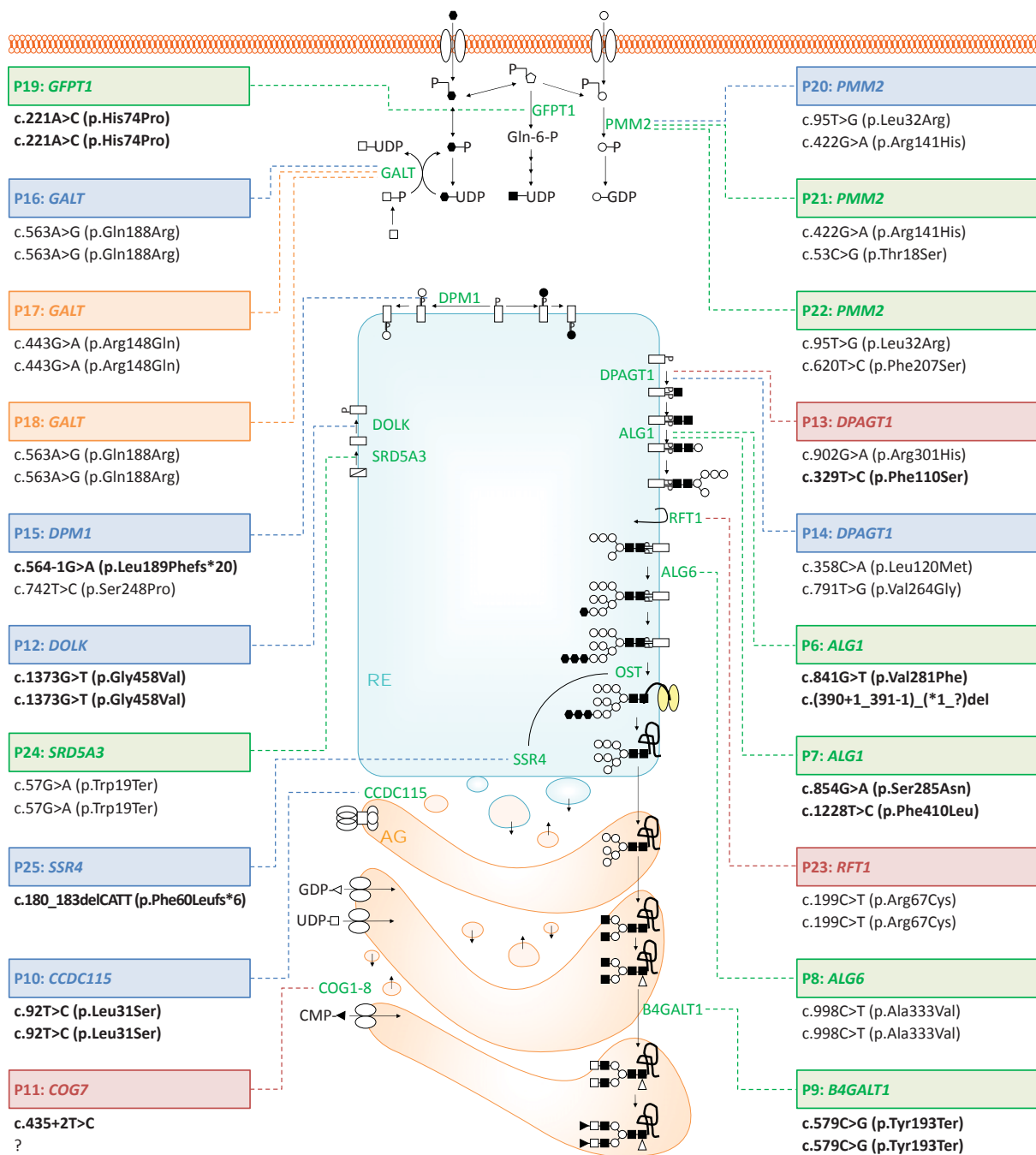
### 3.1.2 Espectro genético y mutacional de pacientes CDG

Del conjunto de 39 pacientes seleccionados por presentar una clínica y/o un perfil bioquímico compatible con un defecto en la ruta de glicosilación de proteínas, se han detectado 20 de ellos que portaban mutaciones en 14 genes asociados a CDG (**Figura 19 y 20**). Se han identificado dos pacientes deficientes en *ALG1* ( $\beta$ -1,4-manosiltransferasa, MIM 605907), uno en *ALG6* ( $\alpha$ -1,3-glucosiltransferasa, MIM 604566), uno en *B4GALT1* ( $\beta$ -1,4-galactosil transferasa 1, MIM 137060), uno en *CCDC115* (factor de ensamblaje de la ATPasa vacuolar, MIM 613734), otro en *COG7* (subunidad 7 del complejo oligomérico de Golgi, MIM 606978), otro en *DOLK* (dolicol quinasa, MIM 610746), dos en *DPAGT1* (dolicol-fosfato N-acetilglucosaminafosfotransferasa 1, MIM 191350), uno en *DPM1* (dolicol-fosfato manosiltransferasa 1, MIM 603503), tres en *GALT* (galatosa-1-fosfato uridiltransferasa, MIM 606999), uno en *GFPT1* (glutamina-fructosa-6-fosfato transaminasa 1, MIM 138292), tres en el *PMM2* (fosfomanomutasa 2, MIM 601785), un paciente de *RFT1* (flipasa, MIM 611908), uno de *SRD5A3* (esteroide 5- $\alpha$ -reductasa 3, MIM 611715) y uno de *SSR4* (subunidad del complejo TRAPD, MIM 300090).

Todos los genes en los que se han encontrado mutaciones están implicados en la ruta de síntesis del glicano unido a dolicol (CDG-I), excepto *B4GALT1*, *COG7* y *CCDC115* que son CDG-II. *B4GALT1* codifica para una galactosil transferasa que participa en el procesamiento del glicano una vez unido a la proteína. *COG7* y *CCDC115* están implicados en la homeostasis del AG manteniendo la correcta distribución de proteínas en las distintas cisternas del Golgi o participando en el ensamblaje de una bomba de protones encargada de mantener el pH del Golgi, respectivamente.



**Fig 19. Espectro genético de los pacientes CDG.** Se han identificado dos pacientes deficientes en *ALG1*, otros dos en *DPAGT1*, tres en *PMM2* y otros tres en *GALT*. Para los demás genes sólo se ha detectado un paciente.

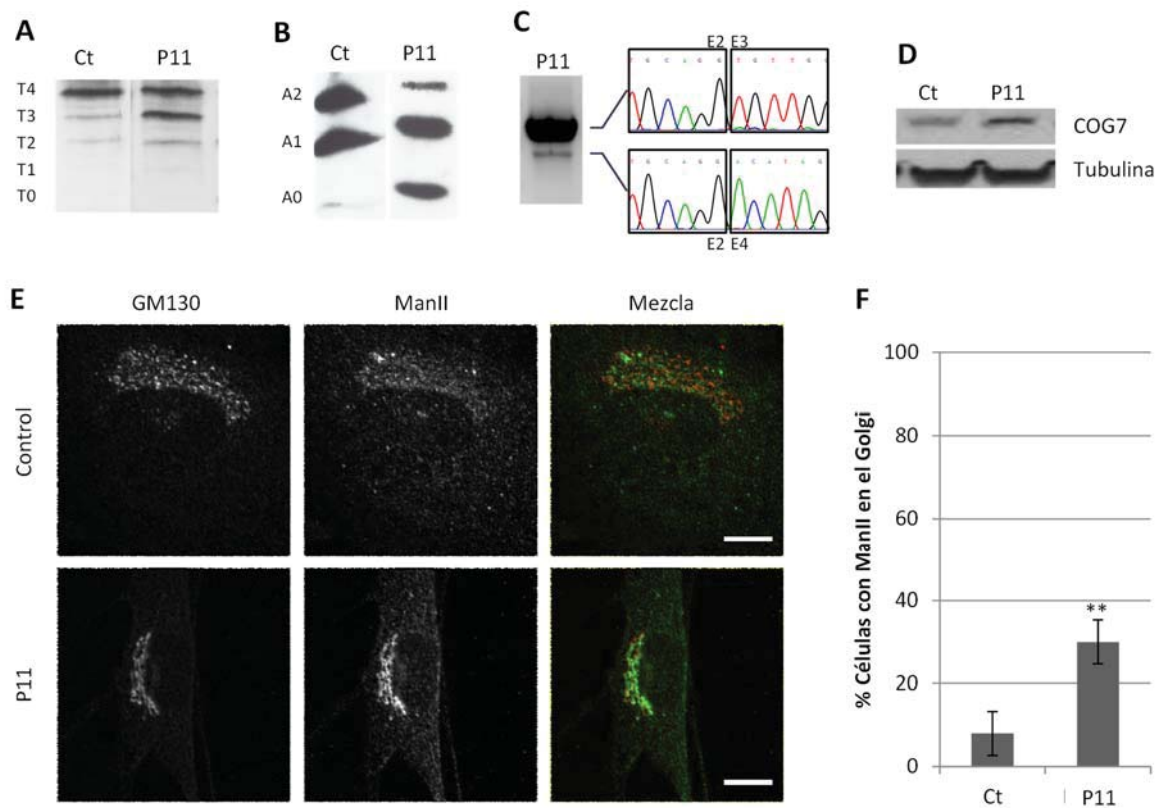


**Fig 20.** Mutaciones identificadas en los genes implicados en la N-glicosilación de proteínas. Los colores de los encabezados indican la aproximación mediante la cual se han identificado (Sanger, TES, CES o WES). Las mutaciones nuevas se resaltan en negrita.

◊ fructosa, ○ manosa, ● glucosa, △ fucosa, ▲ ácido siálico, □ galactosa, ■ N-acetilglucosamina, Gln glucosamina, □ dolicol, ▭ poliprenol.

Resumiendo, se han identificado 11 mutaciones asociadas a CDG y 12 cambios nuevos (**Tabla 15**). Las variantes nuevas incluyen una delección en *SSR4* (c.180\_183delCATT) que presumiblemente produciría una variante de tipo *frameshift*, otra delección en *ALG1* (c.(390+1\_391-1)(\*1\_?)del), un cambio intrónico en *DPM1* (c.564-1G>A) y otro en *COG7* (c.435+2T>C) que podrían afectar al *splicing* y un total de ocho cambios nucleotídicos puntuales que podrían producir un cambio *nonsense* (c.579C>G en *B4GALT1*) o cambios *missense* (c.841G>T, c.854G>A y c.1228T>C en *ALG1*, c.92T>C en *CCDC115*, c.1373G>T en *DOLK* y c.329T>C en *DPAGT1*) o afectar al *splicing* (c.221A>C en *GFPT1*).

En el caso del gen *COG7*, se ha detectado una única variante (c.435+2T>C) presumiblemente patogénica en heterocigosis en el gen *COG7* (subunidad 7 del complejo COG, MIM 606978) en el P11 mediante TES-CDG. Su reanálisis por Sanger, WES y CES no ha permitido la identificación de una segunda variante en este mismo gen. El análisis de la profundidad de cobertura descartó la presencia de grandes delecciones en *COG7*. El P11 había sido clasificado como CDG-II con un posible defecto combinado en la N- y O- glicosilación de proteínas por presentar un patrón tipo 2 en el Trf-IEF y un perfil tipo ApoCIII-0 en el ApoCIII-IEF (**Figura 21 A y B**). El análisis del perfil transcripcional mostró un alelo con un *skipping* en fase del exón 3 y otro alelo completo, cuya secuenciación por Sanger convencional mostró una secuencia normal o *wild-type* (**Figura 21 C**). También se analizó la expresión de la proteína COG7 por *western-blot*, mostrando un resultado comparable a los controles (**Figura 21 D**). Debido a la implicación de la proteína COG7 en el transporte vesicular retrógrado del AG al RE, se trataron los fibroblastos derivados del paciente con Brefeldin A (BFA), un metabolito fúngico que altera la estructura y función del AG induciendo la redistribución de las enzimas del Golgi, como Manosidasa II (ManII), al RE vía transporte retrógrado. Tras el tratamiento con BFA las células control mostraron una localización de ManII en el RE, mientras que un 30% de las células derivadas del P11 mostraron una localización de ManII en el AG, lo que parece indicar un defecto en el transporte retrógrado en las células del P11 (**Figura 21 E y F**).



**Fig 21. Análisis de biomarcadores y estudios moleculares del P11.** (A) El Trf-IEF de suero del P11 muestra una elevación de tri- y disialotransferrina. (B) El ApoCIII-IEF muestra un aumento de los niveles de mono- y asialoApoCIII en el suero de P11. (C) El análisis del cDNA de *COG7* del P11 muestra un alelo *wild-type* y un alelo con un *skipping* en fase del exón 3. (D) Niveles de expresión de la proteína COG7 en unos fibroblastos control (Ct) y del P11. (E) Evaluación del transporte retrógrado AG-RE por inmunofluorescencia de los fibroblastos derivados de un control y del P11 tratados con BFA y teñidos con anti-ManII (verde) y anti-GM130 como marcador del AG (rojo) y las imágenes superpuestas. Escala 10µm. (F) Porcentaje de células con el AG teñido con anti-ManII. El resultado es un promedio de tres experimentos diferentes. Las barras de error indican la desviación estándar. \*\*  $p \leq 0,01$ . T, transferrina; A, ApoCIII y el número que les sigue indica el número de residuos de ácido siálico.

Tabla 15. Resumen de datos clínicos, bioquímicos y genéticos de los pacientes CDG.

Paciente	Sexo <sup>1</sup>	Edad al debut / Datos clínicos <sup>2</sup>	Datos bioquímicos <sup>3</sup>	Gen	Genotipo <sup>4</sup>	
					Alelo paterno	Alelo materno
P6	F	8a. Retraso psicomotor, epilepsia, hipotonía, convulsiones, encefalopatía, dismorfias.	Elevado T0 y T2.	ALG1	c.(390+1_391-1)_(*1_?)del	c.841G>T p.Val281Phe
P7 <sup>5</sup>	F	1a. Retraso psicomotor, microcefalia, convulsiones, escoliosis.	Elevado T0 y T2.	ALG1	c.854G>A p.Ser285Asn	c.1228T>C p.Phe410Leu
P8	F	3a. Retraso psicomotor, hipotonía, macroglosia, ataxia, convulsiones.	Elevado T0 y T2.	ALG6	c.998C>T p.Ala333Val	c.998C>T p.Ala333Val
P9	F	4m. Dismorfias, mamilas invertidas, coagulopatía, hepatomegalia.	Elevado T0, T1, T2 y T3. Elevado AST, CK y TSH.	B4GALT1	c.579C>G p.Tyr193Ter	c.579C>G p.Tyr193Ter
P10	M	2a Hepatopatía, distribución anómala de la grasa, problemas en el aprendizaje, acumulación de cobre en el hígado.	Elevado T2 y T3. Elevado A1. Elevado AST, ALT y colesterol. Transporte retrógrado normal.	CCDC115	c.92T>C p.Leu31Ser	c.92T>C p.Leu31Ser
P11	F	2a. Hipotonía, retraso psicomotor, microcefalia, dismorfias, atrofia subcortical.	Elevado T3. Elevado A0. Transporte retrógrado alterado.	COG7	?	c.435 + 2T>C
P12	M	1a. Sin datos. Fallecido.	Elevado T0 y T2.	DOLK	c.1373G>T p.Gly458Val	c.1373G>T p.Gly458Val
P13	M	4a. Hipotonía, sordera neurosensorial, miopatía, retraso en motor y del lenguaje.	Elevado T0 y T2.	DPAGT1	c.902G>A p.Arg301His	c.329T>C p.Phe110Ser
P14	F	1a. Hipotonía, miopatía, síndrome congénito miasténico.	Elevado T0 y T2.	DPAGT1	c.791T>G p.Val264Gly	c.358C>A p.Leu120Met
P15	F	2m. Convulsiones, cardiopatía, hepatomegalia, hipotonía.	Elevado T0 y T2. Elevado AST y ALT.	DPM1	c.742T>C p.Ser248Pro	c.564-1G>A p.Leu189Phefs*2 0



Tabla 15 (continuación). Resumen de datos clínicos, bioquímicos y genéticos de los pacientes CDG.

Paciente	Sexo <sup>1</sup>	Edad al debut / Datos clínicos <sup>2</sup>	Datos bioquímicos <sup>3</sup>	Gen	Genotipo <sup>4</sup>	
					Alelo paterno	Alelo materno
P16	M	2m. Hipotonía, neuropatía, coagulopatía, hepatopatía.	Elevado T0 y T2.	GALT	c.563A>G p.Gln188Arg	c.563A>G p.Gln188Arg
P17	M	1m. Hipotonía, coagulopatía, hepatopatía, hipotiroidismo, vómitos.	Elevado T0 y T2.	GALT	c.443G>A p.Arg148Gln	c.443G>A p.Arg148Gln
P18	M	1m. Coagulopatía, hipotonía, encefalopatía, hepatopatía.	Elevado T0 y T2.	GALT	c.563A>G p.Gln188Arg	c.563A>G p.Gln188Arg
P19	F	3a. Retraso en el crecimiento, hipotonía, retraso psicomotor, movimientos anómalos de los ojos.	Elevado T0 y T2.	GFPT1	<b>c.221A&gt;C p.His74Pro</b>	<b>c.221A&gt;C p.His74Pro</b>
P20	M	4a. Retraso psicomotor, hipotonía, dismorfias, ataxia.	Elevado T0 y T2.	PMM2	c.95T>G p.Leu32Arg	c.620T>C p.Phe207Ser
P21	M	1a. Coagulopatía, hipotonía, vómitos, convulsiones, retraso en el desarrollo.	Elevado T0 y T2.	PMM2	c.422G>A p.Arg141His	c.53C>G p.Thr18Ser
P22	M	2a. Hepatopatía, retraso motor y del lenguaje, afectación neuronal.	Elevado T0 y T2.	PMM2	c.422G>A p.Arg141His	c.95T>G p.Leu32Arg
P23	F	4m. Convulsiones, dismorfias, hipotonía, sordera neurosensorial, cardiopatía, retraso en el desarrollo.	Elevado T0 y T2.	RFT1	c.199C>T p.Arg67Cys	c.199C>T p.Arg67Cys
P24	M	4a. Retraso en el desarrollo, problemas oculares, hipoplasia cerebelosa.	Elevado T2.	SRD5A3	C.57G>A p.Trp19Ter	c.57G>A p.Trp19Ter
P25	M	2a. Retraso en el desarrollo, autista, dismorfias, hipotonía, episodios febriles, encefalopatía.	Elevado T2.	SSR4	<sup>(6)</sup> <b>c.180_183delCATT p.Phe60Leufs* 6-</b>	

<sup>1</sup> M, masculino; F, femenino. <sup>2</sup> a, años; m, meses. <sup>3</sup> T indica transferrina, A apolipoproteína-CIII y el número que le sigue hace referencia al número de ácido siálico unidos. <sup>4</sup> La nomenclatura de las mutaciones se realizó sobre el transcrito de referencia (**Anexo 3**) siguiendo la normativa establecida por HGVS. Las mutaciones no descritas hasta la fecha se resaltan en negrita. <sup>5</sup> Estudio de segregación de las mutaciones en la familia no realizado por falta de muestra de los padres. <sup>6</sup> Mutación de novo.

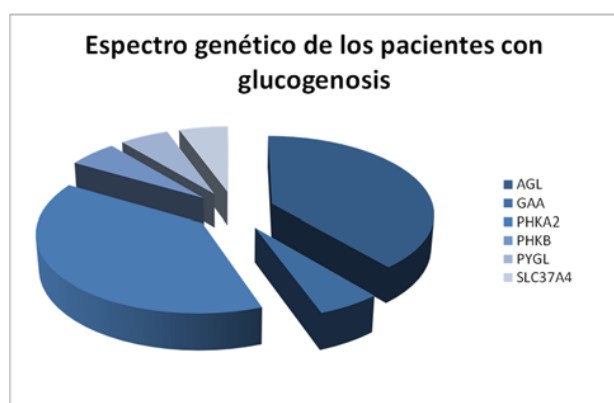


### 3.2 Pacientes con glucogenosis. Espectro genético y mutacional

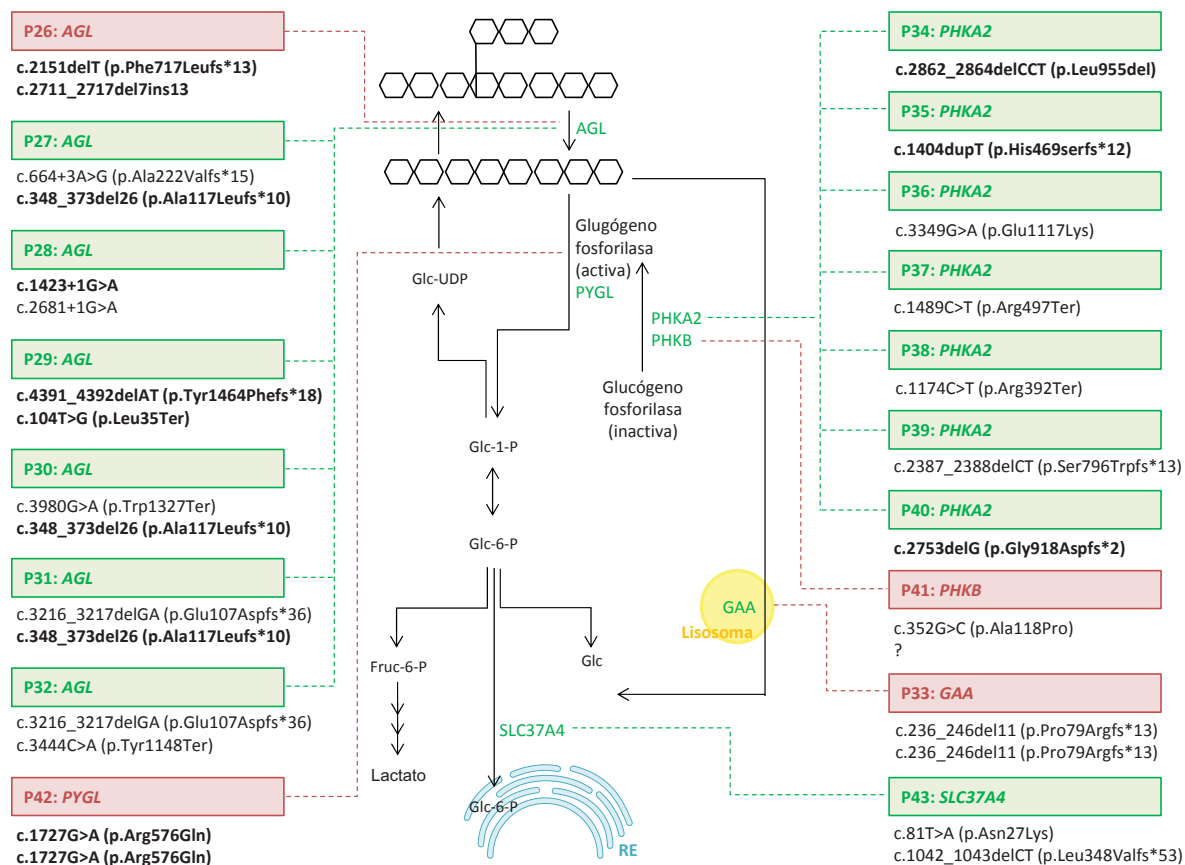
Del conjunto de 47 pacientes seleccionados con sospecha de glucogenosis, se han caracterizado 18 de ellos que portaban mutaciones en seis genes asociados a glucogenosis (**Figuras 22 y 23**). Se ha detectado un paciente con deficiencia en *PYGL* (glucógeno fosforilasa hepática, MIM 613741), otro en *GAA* ( $\alpha$ -1,4-glucosidasa ácida lisosomal, MIM 606800), otro en *SLC37A4* (glucosa-6-fosfato translocasa, MIM 602671), siete en *AGL* (amilo  $\alpha$ -1,6-glucosidasa, MIM 610860) y otros siete en *PHKA2* (subunidad  $\alpha$  de la fosforilasa kinasa hepática, MIM 300798). Además, se ha detectado otro paciente que portaba en heterocigosis una mutación descrita en el gen *PHKB* (subunidad  $\beta$  de la fosforilasa kinasa, MIM 172490).

Todos los genes en los que se han detectado mutaciones codifican para proteínas implicadas en la ruta de degradación del glucógeno excepto *SLC37A4* que está implicado en la homeostasis celular de la glucosa.

En resumen, se han identificado 13 mutaciones asociadas a glucogenosis y 10 variantes nuevas presumiblemente patogénicas. Las variantes nuevas incluyen tres pequeñas deleciones en *AGL* (c.348\_373del26, c.2151delT y c.4391\_4392delAT) y otras dos en *PHKA2* (c.1404dupT y c.2753delG) que posiblemente produzcan un *frameshift* en la proteína, una variante puntual en *AGL* (c.3980G>A) que produciría un cambio *nonsense*, una variante intrónica en *AGL* (c.1423+1G>A) que posiblemente afecte al *splicing*, un cambio nucleotídico en *PYGL* (c.1727G>A) que podría producir un cambio *missense*, una deleción en fase en *PHKA2* (c.2862\_2864delCCT) y una deleción-inserción en *AGL* (c.2711\_2717del7ins13) (**Tabla 16**).



**Fig 22.** Espectro genético de los pacientes con glucogenosis. Se han identificado siete pacientes deficientes en *AGL* y otros siete en *PHKA2*. Para los demás genes sólo se ha detectado un paciente.



**Fig 23.** Mutaciones identificadas en los genes implicados en el metabolismo del glucógeno. Los colores de los encabezados indican la aproximación mediante la cual se han identificado (TES o CES). Las mutaciones nuevas se resaltan en negrita. ○○○ glucógeno, Fruc fructosa, Glc glucosa, P fosfato.

Tabla 16. Resumen de datos clínicos y genéticos de los pacientes con glucogenosis.

Paciente	Sexo <sup>1</sup>	Edad al debut / Datos clínicos <sup>2</sup>	Gen	Genotipo <sup>3</sup>	
				Alelo paterno	Alelo materno
P26	F	8m. Sin datos.	AGL	c.2151delT p.Phe717Leufs*13	c.2711_2717del7ins13 p.(Leu904_Gln906)delinsPKDLM)
P27	F	6a. Sin datos.	AGL	c.664+3A>G p.Ala222Valfs*15	c.348_373del26 p.Ala117Leufs*10
P28	F	8m. Hepatomegalia, hipoglucemia, cardiomiopatía, elevados niveles de CK.	AGL	c.1423+1G>A	c.2681+1G>A
P29	F	15m. Hepatomegalia, retraso en el crecimiento, hipertransaminasemia, episodios de hipoglucemia, hipertrigliceridemia.	AGL	c.4391_4392delAT p.Tyr1464Phefs*18	c.104T>G p. Leu35Ter
P30	M	11m. Miopatía, hipoglucemia, hipertransaminasemia, aumento de CK.	AGL	c.3980G>A p.Trp1327Ter	c.348_373del26 p.Ala117Leufs*10
P31 <sup>4</sup>	F	25a. Sin datos.	AGL	c.3216_3217delGA p.Glu1072Aspfs*36	c.348_373del26 p.Ala117Leufs*10
P32	F	9m. Hepatomegalia, hipertransaminasemia, hipoglucemia, hipertrigliceridemia, aumento de CK y LDH.	AGL	c.3216_3217delGA p.Glu1072Aspfs*36	c.3444C>A p.Tyr1148Ter (CM125030)
P33	F	3m. Cardiomiopatía hipertrófica.	GAA	c.236_246del11 p.Pro79Argfs*13	c.236_246del11 p.Pro79Argfs*13

Tabla 16 (continuación). Resumen de datos clínicos y genéticos de los pacientes con glucogenosis.

Paciente	Sexo <sup>1</sup>	Edad al debut / Datos clínicos <sup>2</sup>	Gen	Genotipo <sup>3</sup>	
				Alelo paterno	Alelo materno
P34	M	4a. Hepatomegalia leve, microesteatosis, hipertransaminasemia.	PHKA2	-	c.2862_2864delCCT p.Leu955del
P35	M	1a. Hepatomegalia, hipertransaminasemia, ferropenia, deficiencia de vit D., hemoglobinopatía.	PHKA2	-	c.1404dupT p.His469Serfs*12
P36	M	15m. Hepatoesplenomegalia, hipertransaminasemia, hipoglucemia.	PHKA2	-	c.3349G>A p.Glu1117Lys
P37	M	10m. Hepatomegalia, hipoglucemia, hipertransaminasemia, deficiencia de vit D.	PHKA2	-	c.1489C>T p.Arg497Ter
P38	M	<2a. Hepatomegalia, acumulación de glucógeno en hígado, hipoglucemia, hipertransaminasemia.	PHKA2	-	c.1174C>T p.Arg392Ter
P38	M	Infancia. Hepatomegalia, hipoglucemia.	PHKA2	-	c.2387_2388delCT p.Ser796Trpfs*13
P40	M	5a. Hepatomegalia leve, hipertransaminasemia.	PHKA2	-	c.2753delG p.Gly918Aspfs*2
P41	F	1m. Hepatomegalia leve, hipertransaminasemia e hipoglucemia.	PHKB	p.Ala118Pro	?
P42	M	38a. Astenia, hipoglucemia y fatiga tras el ejercicio, acumulación de glucógeno en músculo.	PYGL	c.1727G>A p.Arg576Gln	c.1727G>A p.Arg576Gln
P43	M	Infancia. Hepatomegalia, facciones dismórficas, estatura baja, fallo renal, hipoglucemia, acidosis láctica, hiperuricemia, hiperlipidemia.	SLC37A4	c.81T>A p.Asn27Lys	c.1042_1043delCT p.Leu348Valfs*53

<sup>1</sup> M, masculino; F, femenino. <sup>2</sup> a, años; m, meses. <sup>3</sup> La nomenclatura de las mutaciones se realizó sobre el transcrito de referencia (**Anexo 4**) siguiendo la normativa establecida por HGVS. Las mutaciones no descritas hasta la fecha se resaltan en negrita. <sup>4</sup> Estudio de segregación de las mutaciones en la familia no realizado por falta de muestra de los padres.

### 3.3 Hallazgos inesperados

Entre el conjunto de pacientes con sospecha de CDG se ha detectado un paciente que porta variantes posiblemente patogénicas en un gen no asociado a los CDG. Asimismo, entre el conjunto de pacientes con sospecha de glucogenosis se han detectado mutaciones en otros cinco genes no asociados a glucogenosis (*ALDOB*, *ANO5*, *CPT2*, *LIPA*, y *NKX2-5*) que producen características fenotípicas solapantes. (Tabla 17).

Se detectaron dos variantes no descritas presumiblemente patogénicas en el P44 en el gen *FOLR1*. Las variantes nuevas son dos cambios nucleotídicos que podrían producir un cambio *nonsense* (c.321C>A) y otro *missense* (c.622C>T). El cambio *missense* fue clasificado como patogénico por todos los programas de predicción *in-silico* utilizado. *FOLR1* codifica para el receptor de folato cerebral y su deficiencia (MIM 136430) produce afectación neurológica. El análisis bioquímico realizado posteriormente mostró niveles no detectables de 5-metiltetrahidrofolato en el líquido cefalorraquídeo del paciente (datos no mostrados), lo que apoya la deficiencia en *FOLR1*.

Se ha detectado la mutación más frecuente en el gen *ALDOB* en el P45 (c.448G>C). Este gen codifica para fructosa-1,6-bifosfato aldolasa B, que cataliza la conversión de fructosa-1,6-bifosfato a dihidroxiacetona fosfato y gliceraldehído-3-fosfato. Mutaciones en este gen producen intolerancia a la fructosa hereditaria (MIM 612724) y causan problemas gastrointestinales y hepáticos.

También se ha detectado la mutación más frecuente en el gen *ANO5* en el P46 (c.191dupA). Las mutaciones en *ANO5* (MIM 608662) han sido asociadas a distrofia muscular de extremidades y cintura.

El P47 porta una mutación *missense* en homocigosis previamente descrita en el gen *CPT2* (c.338C>T). *CPT2* codifica el enzima carnitina palmitoil transferasa II (MIM 600650) implicada en la ruta de la beta-oxidación de ácidos grasos.

En el P48 se han detectado dos mutaciones previamente descritas en el gen *LIPA*, que codifica para la lipasa lisosomal ácida (LAL). La mutación c.894G>A afecta al proceso de *splicing* (Pagani et al., 1998) y se asocia frecuentemente con la enfermedad de acúmulo de ésteres de colesterol, una forma leve de la deficiencia en LAL, mientras que c.398delC está asociada al fenotipo severo conocido como enfermedad de Wolman. La deficiencia en LAL (MIM 613497) fue confirmada posteriormente por ensayo enzimático.

Además, se ha detectado una mutación en homocigosis en el gen *NKX2-5* en el P49 (c.73C>T). Mutaciones en *NKX2-5* (MIM 600584) se han asociado a problemas cardíacos.

Tabla 17. Datos clínicos y espectro mutacional de pacientes que portan mutaciones en genes no asociados a CDG o glucogenosis.

Paciente	Sexo <sup>1</sup>	Edad al debut / Datos clínicos <sup>2</sup>	Sospecha clínica	Gen Transcrito	Genotipo <sup>3</sup>	
					Alelo paterno	Alelo materno
P44	M	6a. Miocardiopatía hipertrófica, retraso psicomotor, epilepsia mioclónica, leucodistrofia.	CDG	FOLR1 NM_016725	<b>c.321C&gt;A</b> <b>p.Tyr107Ter</b>	<b>c.622C&gt;T</b> <b>p.Arg208Cys</b>
P45	F	1a. Hepatomegalia, retraso en el crecimiento.	GSD	ALDOB NM_000035	c.448G>C p.Ala150Pro	c.448G>C p.Ala150Pro
P46	M	32a. Vacuolas musculares de glucógeno, CK elevada, debilidad muscular	GSD	ANO5 NM_213599	c.191dupA p.Asn64Lysfs*15	c.191dupA p.Asn64Lysfs*15
P47	M	13a. Rabdomiolisis, dolor muscular, hepatomegalia.	GSD	CPT2 NM_000098	c.338C>T p.Ser113Leu	c.338C>T p.Ser113Leu
P48	F	12a. Parénquima hepático desestructurado, microesteatosis hepática, fibrosis hepática.	GSD	LIPA NM_000235	c.398delC p.Ser133Ter	c.894G>A p.Ser275_Gln298del
P49	M	7a. Cardiomiopatía hipertrófica biventricular, derrame pericárdico, sordera neurosensorial, retraso psicomotor, necrosis tubular renal. Fallecido (7a)	GSD	NKX2-5 NM_004387	c.73C>T p.Arg25Cys	c.73C>T p.Arg25Cys

<sup>1</sup> M, masculino; F, femenino. <sup>2</sup> a, años; m, meses. <sup>3</sup> La nomenclatura de las mutaciones se realizó sobre el transcrito de referencia siguiendo la normativa establecida por HGVS. Las mutaciones no descritas hasta la fecha se resaltan en negrita.

# DISCUSIÓN

1.	EVALUACIÓN DE DISTINTAS APROXIMACIONES DE SECUENCIACIÓN MASIVA	93
2.	APLICACIÓN DE LA NGS AL DIAGNÓSTICO GENÉTICO DE PACIENTES	98
2.1	Espectro genético y mutacional de pacientes CDG	98
2.2	Espectro genético y mutacional de pacientes de glucogenosis	102
2.3	Hallazgos inesperados	103

## 1. EVALUACIÓN DE LAS DISTINTAS APROXIMACIONES DE SECUENCIACIÓN MASIVA

---

La tecnología de secuenciación masiva ha supuesto una revolución técnica en el campo del diagnóstico genético al abrir nuevas posibilidades en la identificación de genes implicados en patología en humanos (Koboldt et al., 2013). Este trabajo refleja la aplicación de una tecnología vanguardista para la detección y caracterización de mutaciones en dos grupos de enfermedades genéticamente heterogéneas, los CDG y la glucogenosis.

Durante el desarrollo de esta tesis hemos aplicado tres aproximaciones de secuenciación masiva, un panel personalizado de genes candidatos asociados a la enfermedad en estudio (TES-CDG y TES-EMH), el exoma completo (WES) y un panel comercial de Illumina® que incluye todos los genes asociados a enfermedad en humanos y anotados en la base de datos del OMIM-2013 (CES).

En primer lugar, utilizamos dos paneles personalizados diseñados en el laboratorio donde el número de genes incluidos estaba limitado a aquellos en los que existía suficiente evidencia científica de su relación con la enfermedad a estudio. Actualmente, la utilización de paneles de genes acotados sigue siendo recomendada por la *American College of Medical Genetics and Genomics* (ACMG) ya que tienen buenas coberturas y los hallazgos inesperados no relacionados con la patología son menos numerosos (Rehm et al., 2013). Así, hemos empleado un panel que incluía 43 genes CDG (TES-CDG) y otro que capturaba la región exómica de 111 genes asociados a metabolopatías, entre ellos, 22 de glucogenosis (TES-EMH). La tasa de diagnóstico del panel TES-CDG fue baja (19%). Una de las posibles causas podría ser el limitado número de genes incluidos en el diseño inicial. Los CDG son un grupo de enfermedades emergentes donde nuevos genes son descritos con regularidad (Matthijs et al., 2012). Así, desde el diseño de nuestro panel al inicio de esta tesis, quince genes nuevos han sido asociados a los CDG (Carss et al., 2013; Cossins et al., 2013; Foulquier et al., 2012; Jansen et al., 2016a; Jansen et al., 2016b; Jansen et al., 2016c; Koehler et al., 2013; Losfeld et al., 2014; Matalonga et al., 2016; Ng et al., 2013; Park et al., 2015; Pérez et al., 2012; Rafiq et al., 2011; Senderek et al., 2011; Timal et al., 2012). Esta limitación del panel TES-CDG ha quedado puesta de manifiesto tras identificar mediante CES o WES dos pacientes CDG deficientes en *SSR4* o *GFPT1* que habíamos analizado previamente por TES-CDG sin llegar a un diagnóstico genético ya que estos genes no estaban incluidos en el diseño. Así pues, mientras no esté definido por completo el espectro genético de los CDG, el diagnóstico mediante paneles de genes personalizado es poco adecuado para este grupo de enfermedades.

Otra posible explicación al bajo rendimiento de nuestros paneles puede ser que tanto los CDG como la glucogenosis son enfermedades que debutan, en general, con una clínica poco específica y carecen de un biomarcador patognomónico para su identificación, por lo que cabe la posibilidad de



que alguno de los pacientes seleccionados estén afectados por otra enfermedad diferente a la que fueron dirigidos inicialmente.

Por este motivo y con el fin de mejorar el diagnóstico genético y poder resolver los casos no resueltos, comenzamos a utilizar el análisis del exoma completo (WES) que, al capturar la región exómica de los 22.000 genes que codifican para proteínas, permite un análisis no sesgado de genes. De esta manera, hemos identificado un paciente deficiente en *CCDC115*, un gen nuevo no vinculado a ninguna patología hasta la fecha. Cabe destacar que, aunque esta aproximación es adecuada para identificar genes nuevos, su aplicación en la práctica clínica rutinaria resulta comprometida porque no tiene ni cobertura horizontal completa ni profundidad de cobertura homogénea (Choi et al., 2009). Por ejemplo, en nuestro caso, no fuimos capaces de detectar por WES la mutación c.422G>A en *PMM2* en uno de los pacientes CDG. Aunque la profundidad de cobertura media de ese gen para ese paciente era >20x, c.422 estaba representado sólo por cuatro lecturas, por lo que fue excluido durante el filtrado de variantes al no tener una DC>10x. Además, sólo una de las cuatro lecturas contenía la mutación por lo que, aun permitiendo un filtrado más laxo con variantes representadas por menos lecturas, los parámetros de calidad del *variant calling* habrían seguido descartando este cambio por ser un *strand bias*. La tecnología ha mejorado sustancialmente desde su aplicación en el presente trabajo, lo que permitirá la utilización de forma más eficaz de esta aproximación en la práctica clínica habitual (Tarailo-Graovac et al., 2016).

El panel comercial de genes asociados a enfermedad anotados en la base de datos del OMIM (CES) es actualmente una buena aproximación a emplear para el diagnóstico de enfermedades genéticamente muy heterogéneas como los CDG o la glucogenosis. CES proporciona una profundidad de cobertura mayor y más uniforme que WES, a la vez que permite un estudio más amplio en genes no asociados inicialmente a la patología a estudio. Así, mediante esta aproximación, hemos detectado mutaciones en *FOLR1* entre el grupo de pacientes con sospecha de CDG y en *ANO5*, *CPT2*, *LIPA* o *NKX2-5* entre aquellos con sospecha de glucogenosis. Esto puede ser debido a que estos pacientes presentaban un fenotipo similar a las enfermedades a estudio, o bien a que debutaron con una presentación atípica de la enfermedad.

NGS, además de permitir el análisis simultáneo de varios genes, resulta ventajoso frente a Sanger en genes largos como *AGL*, ya que al poder secuenciar sus 34 exones en una única carrera se reduce el tiempo y el coste del análisis. Asimismo, se evita la amplificación recurrente y exclusiva de uno de los alelos (*allele dropout*). En nuestro caso, dos variantes identificadas por NGS en heterocigosis en los genes *AGL* y *ALG1* fueron detectadas en homocigosis mediante la secuenciación de Sanger. Una de las posibles causas del *allele dropout* es el uso de oligonucleótidos complementarios a regiones donde hay SNP que provocarían una hibridación incompleta de los mismos y, como consecuencia, una falta de amplificación del segmento

flanqueado (Robyns et al., 2015; Zhang et al., 2015). La metodología subyacente a la NGS empleada en este trabajo, en la que la amplificación se realiza de manera clonal utilizando oligonucleótidos universales que hibridan con los adaptadores previamente unidos al DNA, elimina esta posibilidad. Sin embargo, a partir de los archivos BAM de estos pacientes analizamos las regiones donde hibridaban los oligonucleótidos utilizados en Sanger, sin detectar ningún SNP que pudiese explicar la amplificación preferente de uno de los alelos. En nuestros casos, es posible que la causa del *allele dropout* se deba a la dificultad de la polimerasa de acceder a determinadas regiones de la cromatina (Blais et al., 2015), problema que es evitado en NGS al partir de una muestra fragmentada al azar.

Además de lo indicado anteriormente, la NGS permite la detección con precisión de prácticamente todo el espectro mutacional. A partir de los datos de profundidad de cobertura, hemos sido capaces de detectar una delección de parte del gen *ALG1* en el P6. En este caso, puesto que sólo habíamos capturado la región exómica, utilizamos la tecnología de SNP-array para confirmar la delección y, de esta manera comprobamos que tenía una longitud de 50Kb y además abarcaba otro gen completo (*EEF2KMT*).

Sin embargo, a pesar de las ventajas y del gran potencial de la NGS en la identificación de cambios puntuales, carece de suficiente sensibilidad para la detección de algunas variantes estructurales con las que hay que tener especial atención. Se pueden detectar con precisión inserciones/delecciones <20pb con los programas utilizados para el *variant calling* y >100pb cuando se realiza por comparación de la DC, pero existe un rango de imprecisión para la detección de indel de 20~100pb que requiere una inspección visual de las lecturas para su identificación (Feng et al., 2015; Nord et al., 2011). Durante nuestro estudio, tres de los pacientes deficientes en *AGL* presentaron una delección de 26pb y sólo en uno de ellos la variante fue anotada en el VCF durante el *variant calling*, mientras que en los otros dos pacientes, la delección fue detectada por revisión de los archivos BAM de los genes candidatos.

Otro factor importante de la NGS a considerar es la incertidumbre en el alineamiento de lecturas en regiones con alta homología de secuencia. El alineamiento de una lectura en varias regiones del genoma produce una medida de calidad imprecisa que genera secuencias de poca fiabilidad y que finalmente son descartadas durante el análisis bioinformático. Aunque se ha recomendado no analizar por NGS genes que tengan asociados pseudogenes debido a la incapacidad del sistema para determinar si las variantes se encuentran en el gen funcional o en los pseudogenes (Jones et al., 2011), nosotros hemos sido capaces de identificar dos pacientes deficientes en *ALG1*, homólogo en su secuencia a 12 pseudogenes y dos genes codificantes. Cabe resaltar que la detección de las variantes en ambos casos fue mediante CES, y que uno de los pacientes también había sido analizado por TES-CDG y WES sin llegar a un diagnóstico genético. Es posible que la pérdida de las variantes en estas aproximaciones fuera debida al empleo de sistemas más antiguos

de secuenciación que generaban una menor longitud de las lecturas (50pb y 75-100pb respectivamente frente a 250pb de CES), ya que se ha descrito que lecturas más largas aumentan la precisión del alineamiento (Claes and De Leeneer, 2014).

A pesar de la versatilidad y del gran potencial de la NGS, la tasa diagnóstica de ambas enfermedades ha sido cercana al 50%, no completando el diagnóstico en 42 pacientes de un total de 86 analizados. Esto puede ser debido a limitaciones técnicas de la propia metodología empleada, como la falta de sensibilidad para la detección de algunas variantes estructurales o la dificultad en el alineamiento en regiones con alta homología de secuencia, o a limitaciones en el filtrado y la interpretación de los resultados. La adecuación de los filtrados y priorización de los SNV ha de ser acorde al conjunto de enfermedades objeto a estudio (Gilissen et al., 2012). Así pues, teniendo en cuenta la baja incidencia en la población de los CDG y la glucogenosis, cabe esperar que las mutaciones causantes de estas enfermedades tengan una baja frecuencia de segregación y, por lo tanto, una baja frecuencia alélica. Se ha descrito que posiblemente para enfermedades poco frecuentes con un modo de herencia autosómica recesiva, las variantes cuya frecuencia alélica poblacional sea mayor de 0,5% podrían ser consideradas benignas ya que serían demasiado frecuentes para ser patogénicas (Dorschner et al., 2013). Sin embargo, nosotros hemos creído conveniente incrementar este parámetro a 1% porque la frecuencia de portadores de algunas mutaciones es superior a lo esperado, como para la mutación p.Arg141His de *PMM2* que tiene una frecuencia alélica de 0,66% (ExAC).

Otra posible causa de esta tasa diagnóstica intermedia podría ser porque el filtrado de genes candidatos se ha realizado en base a una posible ruta metabólica afectada (glicosilación o síntesis/degradación de glucógeno). Tal y como se ha discutido anteriormente, los CDG y la glucogenosis debutan con un espectro fenotípico amplio que puede solapar con otras enfermedades y un perfil bioquímico que puede ser inespecífico, por lo que cabe la posibilidad de que los pacientes estén afectados por otra patología y sea necesario reevaluar los datos con otros filtros genéticos. En nuestro caso, esta problemática ha quedado ilustrada con la identificación de seis pacientes deficientes en genes no asociados a los CDG o a la glucogenosis.

La selección de variantes candidatas también se apoya en los análisis de patogenicidad obtenidos a partir de los programas informáticos de predicción *in-silico*. Por ello, es importante conocer la base del análisis que realiza cada programa y utilizar varios que empleen diferentes métodos para valorar toda la información en su conjunto. Los programas de predicción *in-silico* que hemos empleado más habitualmente durante el análisis de variantes han sido SIFT, Polyphen2 y *Mutation Taster*. Los dos primeros se basan en la homología de secuencia entre proteínas ortólogas para clasificar los cambios aminoacídicos según su conservación entre especies (Adzhubei et al., 2010; Ng and Henikoff, 2001), mientras que *Mutation Taster* no sólo predice la consecuencia funcional de cambios aminoacídicos sino también de variantes intrónicas, sinónimas e inserciones y/o

deleciones, al integrar la información de otras bases de datos con herramientas de predicción propias (Schwarz et al., 2014). Así pues, los programas de predicción *in-silico* pueden proporcionar un indicio sobre la patogenicidad de un cambio pero esta información no debe ser considerada de manera aislada al tomar una decisión sobre si es causante de enfermedad (Gilissen et al., 2012; Tchernitchko et al., 2004). En este sentido cabe resaltar que la mutación previamente descrita p.Ser248Pro en *DPM1* (García-Silva et al., 2004) detectada en el P15 tiene una predicción *in-silico* de tolerada, benigna y patogénica por SIFT, Polyphen2 y *Mutation Taster*, respectivamente. Esta discrepancia en la predicción de la patogenicidad de esta mutación se debe a una divergencia en la evolución de la estructura de la proteína (Ashida et al., 2006; Maeda et al., 1998), de manera que la mutación genera una predicción de tolerancia en los programas de predicción basados exclusivamente en homología de secuencia como son SIFT y Polyphen2.

También cabe la posibilidad de que los pacientes en los que no hemos llegado a un diagnóstico genético actualmente, posean mutaciones causantes enfermedad en regiones intrónicas no secuenciadas. En principio, parecería poco probable que este fuese el motivo, ya que se ha descrito que el 85% de las mutaciones causantes de enfermedad se encuentran localizadas en los exones (Choi et al., 2009). Sin embargo, puede que este porcentaje esté sobrevalorado, ya que se basa en los análisis exómicos recurrentes de los laboratorios de diagnóstico genético. También es posible que haya variantes causantes de la patología del paciente en regiones no codificantes de proteínas alejadas del gen candidato, tales como regiones de metilación, regiones reguladoras de la transcripción o elementos reguladores de *splicing*, o incluso podrían estar localizadas en genes que se transcriben a RNA no codificante (ncRNA, *non-coding RNA*) (Jarinova and Ekker, 2012). Una alternativa para la identificación de estas variantes sería el estudio del genoma completo (WGS, *whole-genome sequencing*) (Gilissen et al., 2014). Sin embargo, esta aproximación genera muchas variantes de difícil interpretación ya que prácticamente no existen datos de frecuencia alélica en regiones no exónicas. Así, otra buena aproximación diagnóstica para estos casos sería el análisis cuantitativo y cualitativo del transcriptoma celular por RNA-seq (Codina-Sola et al., 2015), incluso en combinación con el genoma completo (Simul-seq) (Reuter et al., 2016).

Este trabajo refleja el continuo y rápido desarrollo de la secuenciación masiva de DNA aplicada al diagnóstico genético, de tal manera que la mejora de esta técnica probablemente la convierta en la metodología de elección en primera instancia, ya que aporta otras ventajas frente a la técnica *gold-standard* utilizada hoy en día, la secuenciación por Sanger convencional, aunque es esencial conocer las ventajas y las limitaciones de cada aproximación para seleccionar la más adecuada en cada caso (O'Donnell-Luria and Miller, 2016; Xue et al., 2015). En resumen, la secuenciación masiva ha supuesto una revolución técnica y conceptual que está modificando incluso la forma de entender la práctica clínica de las enfermedades genéticas humanas, dando paso a la llamada medicina de precisión.

## 2. APLICACIÓN DE LA NGS AL DIAGNÓSTICO GENÉTICO DE PACIENTES \_\_\_\_\_

Hasta hace unos años, el diagnóstico genético se realizaba de forma rutinaria mediante la secuenciación convencional por Sanger tras una serie de aproximaciones clínicas, bioquímicas y/o celulares en un intento de reducir a uno o a un conjunto pequeño de genes candidatos a analizar, que se secuenciaban por orden de frecuencia poblacional. Este proceso resultaba lento y especialmente laborioso en enfermedades genéticamente heterogéneas, quedando numerosos pacientes sin diagnóstico. En este trabajo demostramos la utilidad de la tecnología de la secuenciación masiva en la mejora del diagnóstico genético de enfermedades genéticamente heterogéneas como son los CDG y la glucogenosis.

### 2.1 Espectro genético y mutacional de pacientes CDG

En este trabajo hemos analizado 39 pacientes con características bioquímicas y/o clínicas de CDG (17 CDG-I, 21 CDG-II y uno sin clasificar). Hemos llegado al diagnóstico genético de los 17 pacientes CDG-I (100%), 3 CDG-II (14%) y del paciente sin clasificar, lo que hace un total de 21 (21/39, 54%). Probablemente hemos caracterizado más pacientes CDG-I porque las proteínas implicadas en la ruta de síntesis del oligosacárido estándar unido a dolicol y su transferencia a la proteína naciente son más conocidas. Por el contrario, en la actualidad la ruta de modificación del glicano unido a la proteína no está caracterizada totalmente. Además, recientemente se está ampliando el concepto de CDG al extenderlo a proteínas implicadas en el mantenimiento de la homeostasis del AG, tales como proteínas del tráfico vesicular y transportadores de iones o de azúcares unidos a nucleótidos, que también producirían un impacto en las últimas etapas de la glicosilación (Bexiga and Simpson, 2013; Hennet and Cabalzar, 2015). Por lo que, *a priori*, la identificación de mutaciones en pacientes CDG-II sería más complicado.

De los 21 pacientes en los que hemos llegado a un diagnóstico genético, 19 portaban mutaciones en 13 genes previamente asociados a los CDG (*ALG1*, *ALG6*, *B4GALT1*, *COG7*, *DOLK*, *DPAGT1*, *DPM1*, *GALT*, *GFPT1*, *PMM2*, *RFT1*, *SRD5A3* y *SSR4*), uno portaba mutaciones en un gen no vinculado a ninguna patología hasta la fecha (*CCDC115*) y otro en un gen no asociado a la glicosilación de proteínas (*FOLR1*). En total, hemos identificado 23 mutaciones en 15 genes diferentes, 12 de ellas descritas por primera vez en este trabajo.

La gran heterogeneidad genética de los CDG junto con la dificultad de su identificación clínica se ve reflejada en la detección, en la mayoría de los casos, de un único paciente deficiente para cada gen. La excepción ha sido la deficiencia en *PMM2*, con tres pacientes identificados, tal y como cabía esperar ya que es la forma más frecuente de CDG (Wolfe and Krasnewich, 2013), a pesar de que hemos partido de un conjunto sesgado de pacientes CDG donde la mayoría de los *PMM2*-CDG ya habían sido identificados con anterioridad por Sanger convencional. Además hemos identificado

tres pacientes deficientes en *GALT*, dos en *DPAGT1* y otros dos en *ALG1*. Desde la aplicación de la secuenciación masiva se ha demostrado que la deficiencia en algunos genes ocurre con mayor frecuencia de lo que se había estimado inicialmente, probablemente por presentar heterogeneidad fenotípica (Freeze et al., 2014; Jones et al., 2013; Timal et al., 2012). Esta puede haber sido la situación de *DPAGT1*-CDG, que había sido asociado a un fenotipo multisistémico muy severo (Imtiaz et al., 2012; Wu et al., 2003; Würde et al., 2012) pero, a partir de la detección de pacientes con un fenotipo neuromuscular moderado (Basiri et al., 2013; Belaya et al., 2012; Iqbal et al., 2012) el número de *DPAGT1*-CDG identificados ha aumentado considerablemente.

Además de las mutaciones identificadas en genes CDG descritos, hemos detectado un paciente con mutaciones en un gen no asociado a CDG hasta la fecha (*CCDC115*). El P10 porta un cambio de tipo *missense* (p.Leu31Ser) que afecta a un aminoácido conservado entre especies alejadas filogenéticamente. Este cambio también ha sido detectado en otros siete individuos de cuatro familias no relacionadas que presentaron una sintomatología similar (Jansen et al., 2016b). Es posible que la variante afecte a la función de la proteína ya que se ha comprobado que no afecta a los niveles de proteína. A partir de los estudios funcionales realizados, *CCDC115* parece estar localizado en vesículas de transporte retrógrado AG-RE COP-I, aunque también hemos detectado una colocalización parcial con las vesículas de transporte anterógrado RE-AG COPII y endosomas tempranos. Todavía desconocemos la implicación de *CCDC115* en el transporte vesicular ya que el transporte retrógrado ensayado en fibroblastos tratados con BFA parece no estar afectado. Cabe la posibilidad de que *CCDC115* esté implicado en el correcto ensamblaje de la ATPasa vacuolar (ATPasa-V), ya que se ha visto que su secuencia es homóloga a la proteína Vma22p de levadura. Vma22p está implicada en la estabilización del dominio V0 de la ATPasa-V durante su ensamblaje en el RE. En humanos, se ha descrito la deficiencia en otros dos posibles factores de ensamblaje, en *TMEM199* con síntomas clínicos y hallazgos bioquímicos similares a la deficiencia en *CCDC115* tales como hepatoesplenomegalia, hipertransaminasemia, hipercolesterolemia y baja ceruloplasmina, y en *VMA21* que produce una miopatía ligada al X con autofagia excesiva (XMEA) (Jansen et al., 2016c; Ramachandran et al., 2013). La ATPasa-V está involucrada en el mantenimiento del pH de distintos orgánulos (Forgac, 2007). Cabe esperar que la deficiencia en *CCDC115* afecte al mantenimiento del pH de las cisternas del AG y a su vez a la actividad o localización de algunas glicosiltransferasas (Maeda and Kinoshita, 2010). No obstante, también se ha propuesto una posible implicación del dominio V0 de la ATPasa-V en la fusión de membranas del transporte vesicular (Rosnoblet et al., 2012). Así pues, *CCDC115* parece estar implicado de alguna manera en el mantenimiento de la homeostasis del aparato de Golgi, afectando a la N- y O-glicosilación de proteínas tal y como ha quedado reflejado en los perfiles de isoformas alterados de las proteínas Trf y ApoCIII en el suero del paciente. Además, la estructura dilatada y vacuolizada del RE que hemos observado en los fibroblastos de piel derivados del P10 podría estar indicando una acumulación de glicoproteínas aberrantes, tal y como se ha descrito en otros CDG (Marquardt

et al., 1995). No obstante, sería necesario realizar otros estudios funcionales para elucidar el papel de *CCDC115*, tales como el análisis del transporte vesicular, estudios del pH de los posibles orgánulos implicados o la generación de un modelo celular y/o animal *knock-out* para *CCDC115* utilizando técnicas de edición génica como CRISPR/Cas9.

También hemos detectado un paciente (P11) con un único cambio en heterocigosis (c.435+2T>C, p.Val107\_Gln145del) en *COG7*, un gen CDG con herencia autosómica recesiva. Tanto la predicción funcional como los datos de frecuencia alélica indican que la variante podría ser patogénica, ya que se pierde el 5' *splice site* canónico (gt) produciendo una eliminación en fase del exón 3 y su frecuencia en alelos control es de 0,00014% y nunca ha sido detectada en homocigosis (ExAC, EVS). El P11 presenta una alteración en los perfiles de isoformas de las proteínas N- y O-glicosiladas estudiadas, así como una alteración en el transporte retrógrado de las células tratadas con BFA, lo que parece corroborar una posible deficiencia en las proteínas implicadas en el transporte retrógrado, como son las proteínas que conforman el complejo oligomérico de Golgi (COG), entre las que se encuentra *COG7* (Smith and Lupashin, 2008). Sin embargo, aunque los hallazgos bioquímicos indican que podría tratarse de un paciente deficiente en *COG7*, no hemos encontrado una segunda mutación patogénica en la región exónica analizada y los resultados del perfil transcripcional y de la expresión de proteína sugieren que no habría cambios patogénicos ni en la región intrónica ni en regiones reguladoras de la transcripción. Por otra parte, es de destacar que los glicanos N-unidos a proteínas plasmáticas de P11 fueron normales, lo que contrasta con los estudios en pacientes deficientes en *COG7*, en los que se describe un aumento de glicanos truncados (Guillard et al., 2011a). Asimismo, el fenotipo clínico moderado de P11 también difiere de los casos descritos en la bibliografía donde todos presentaban una clínica muy severa incluso letal en el primer año de vida (Morava et al., 2007; Ng et al., 2007; Wu et al., 2004; Zeevaert et al., 2009). Aunque se ha descrito que cada individuo porta de 10 a 20 variantes de tipo LoF en su exoma (Consortium et al., 2012), resulta llamativa la detección de un portador de una variante posiblemente patogénica en un gen candidato (*COG7*) cuando los hallazgos bioquímicos y celulares son consistentes con una deficiencia en el transporte retrógrado AG-RE. Cabe la posibilidad de que esta variante presumiblemente patogénica en combinación con otras localizadas en otros genes implicados en la misma ruta acaben afectando a la glicosilación a pesar de que no haya una deficiencia completa de ninguno de ellos (sinergismo heterocigoto) (Vockley et al., 2000).

Además de las variantes de *COG7* y de *CCDC115*, hemos identificado otros 10 cambios no descritos hasta ahora; tres variantes de tipo LoF y una variante estructural que *a priori* podrían considerarse patogénicos, y seis cambios puntuales presumiblemente de tipo *missense*. Las variantes LoF comprenden una pequeña delección en *SSR4* (c.180\_183delCATT), una variante de tipo *nonsense* en *B4GALT1* (p.Tyr193Ter) y un cambio que podría afectar al proceso de *splicing* en *DPM1* (c.564-1G>A).



Todos estos cambios de tipo LoF producirían proteínas truncadas carentes de regiones importantes para la función de la correspondiente proteína, bien por afectar a la pauta de lectura (c.180\_183delCATT en *SSR4*), o bien por parar la traducción de forma prematura (p.Tyr193Ter en *B4GALT1*). Las proteínas, si se sintetizasen, carecerían de su extremo C-terminal donde se localizarían los dominios esenciales para la función de la proteína, como son la región transmembrana de anclaje al RE en *SSR4* o los dominios de unión al sustrato y al cofactor en *B4GALT1*.

En el caso del cambio nucleotídico intrónico identificado en *DPM1* (c.564-1G>A), los estudios del perfil transcripcional muestran que la variante produce una eliminación del exón 8 en el proceso de *splicing*, alterando la pauta de lectura. La proteína truncada que se generaría carecería de su extremo carboxi-terminal implicado en la interacción con otra proteína para su anclaje a la membrana del RE (Maeda and Kinoshita, 2008) y donde se localiza uno de los residuos implicado en la interacción con los sustratos (Maeda and Kinoshita, 2008). En resumen, si el transcrito generado no llegase a ser degradado por el sistema *nonsense mediated decay* (NMD), la proteína que se generaría no sería funcional y al estar deslocalizada sería ubiquitinada y finalmente degradada por el proteasoma (Ashida et al., 2006).

Además, hemos detectado una variante estructural que comprende una delección de 50Kb que comienza en el intrón 3 del gen *ALG1* [c.(390+1\_391-1)\_(\*1\_?)del] e incluye el gen *EEF2KMT*. Este último gen codifica una metiltransferasa específica del factor de elongación de la traducción 2 eucariota (eEF2) y su deficiencia no ha sido asociada a ninguna patología hasta la fecha. La presentación clínica del P6 es similar a la de los pacientes ALG1-CDG descritos (Morava et al., 2012), por lo que su fenotipo parece ser debido exclusivamente al defecto de *ALG1*.

En cuanto a los nuevos cambios exónicos identificados que podrían generar mutaciones *missense* (p.Val281Phe, p.Ser285Asn y p.Phe410Leu en *ALG1*, p.Gly458Val en *DOLK* y p.Phe100Ser en *DPAGT1*), podrían considerarse patogénicos en base a su baja frecuencia alélica y/o a la predicción de patogenicidad del cambio aminoacídico. El cambio exónico c.221A>C de *GFPT1* que produciría el cambio aminoacídico p.His74Pro ha sido clasificado como benigno por los programas de predicción *in-silico* basados en homología de secuencia (SIFT y Polyphen2). Este cambio, localizado al final del exón 3, podría afectar al proceso de *splicing* tal y como apunta el *software* de predicción *Mutation Taster*. Para analizar esta posibilidad, sería necesario realizar estudios funcionales *ex vivo* con minigenes (Vega et al., 2009), ya que el estudio del perfil transcripcional no se puede realizar en las muestras accesibles como sangre o fibroblastos porque *GFPT1* no se expresa en estos tejidos.



## 2.2 Espectro genético y mutacional de pacientes de glucogenosis

Durante el desarrollo de esta tesis hemos detectado mutaciones patogénicas en 23 pacientes con sospecha de glucogenosis de un total de 47 pacientes analizados (49%). De ellos, 18 portaban mutaciones bialélicas en genes de glucogenosis excepto uno que presentó un único cambio en heterocigosis. Además, cinco pacientes presentaron mutaciones en otros genes no relacionados con glucogenosis.

Hemos detectado 23 mutaciones en 5 genes: *AGL*, *GAA*, *PHKA2*, *PYGL* y *SLC37A4*. Todos estos genes codifican para proteínas implicadas en la ruta de degradación del glucógeno, excepto *SLC37A4* que está implicado en la homeostasis celular de la glucosa. Dos tercios de los pacientes diagnosticados presentaron mutaciones en *AGL* o *PHKA2* (32% en *AGL* y 32% en *PHKA2*). Esto contrasta con lo publicado por otros autores donde la deficiencia en *PHKA2* es la forma más común de glucogenosis (Ozen, 2007) y con los datos de la Asociación Española de Glucogenosis (AAEEG, [www.glucogenosis.org](http://www.glucogenosis.org)) que refieren las deficiencias en *PYGM* y *GAA* como las más comunes en España. Defectos en estos dos últimos genes causan la enfermedad de McArdle y Pompe, respectivamente, con características clínicas y bioquímicas muy características que hacen que sean más fácilmente reconocibles y diagnosticados, sesgando de esta manera nuestro conjunto de pacientes a estudio.

Además, se ha detectado un paciente heterocigoto para una mutación previamente descrita del gen *PHKB*. Posiblemente se trate de un portador o tenga una segunda mutación en el otro alelo todavía no identificada.

En total, hemos identificado 10 variantes probablemente patogénicas que no habían sido descritas previamente. Seis se encuentran en el gen *AGL*, tres en *PHKA2* y una en *PYGL*. Cinco de las seis variantes que hemos detectado en *AGL* son de tipo LoF. Este tipo de mutaciones constituye casi el 82% de las mutaciones descritas en este gen y anotadas en la base de datos de *HGMD® professional* (Goldstein et al., 2010; Lucchiari et al., 2006; Shen and Chen, 2002). Las variantes LoF nuevas que hemos detectado en *AGL* son de tipo *frameshift* (p.Phe717Leufs\*13, p.Tyr1464Phefs\*18, p.Ala117Leufs\*10), *nonsense* (p.Leu35Ter) y *splicing* (c.1423+1G>A) y generan proteínas carentes de su extremo C-terminal, muy conservado entre vertebrados, donde se encuentran los aminoácidos implicados en la unión al glucógeno (Cheng et al., 2007; Okubo et al., 1999). La sexta variante que hemos identificado en *AGL* es un cambio complejo (c.2711\_2717del7ins13, p.Leu904\_Gln906delinsPKDLM) que elimina tres residuos aminoácidos conservados entre vertebrados, aunque sería necesario realizar estudios funcionales para comprobar la patogenicidad del mismo.

Tres de las mutaciones que hemos identificado en *PHKA2* no habían sido descritas previamente (p.His469Serfs\*12, p.Gly918Aspfs\*2 y p.Leu955del). Las variantes *frameshift* generan una proteína truncada carente de uno o ambos dominios de unión a calmodulina respectivamente y p.Leu955del deleta un aminoácido muy conservado entre especies incluso alejadas filogenéticamente. Si bien, la patogenicidad de la delección en fase habría que comprobarla mediante estudios funcionales.

El cambio *missense* no descrito que hemos detectado en *PYGL* (p.Arg576Gln) ha sido clasificado como probablemente dañino por los programas de predicción *in-silico* utilizados ya que afecta a un aminoácido conservado. Además, este cambio no ha sido detectado en la población general (1000g, EVS, ExAc), lo que sugiere que podría ser causante de enfermedad.

## 2.3 Hallazgos inesperados

La aplicación de CES o WES nos ha permitido detectar mutaciones en otros genes no asociados *a priori* con las enfermedades en estudio y cuya deficiencia produce un fenotipo y/o un espectro bioquímico solapante. De esta manera, hemos identificado un paciente deficiente en *FOLR1* dentro del conjunto de pacientes con sospecha de CDG y cinco pacientes deficientes en *ALDOB*, *ANO5*, *CPT2*, *LIPA* y *NKX2-5* respectivamente entre los pacientes con sospecha de glucogenosis.

El P44 presentaba cardiomiopatía hipertrófica, leucodistrofia, retraso psicomotor, epilepsia mioclónica y un patrón normal de isoformas de la transferrina sérica. Sin embargo, puesto que se han descrito falsos negativos en esta técnica, se incluyó para análisis genético en el conjunto de pacientes con sospecha de CDG. *FOLR1* (MIM 136430) codifica para el transportador de folato cerebral y su deficiencia produce neurodegeneración por una alteración en la síntesis de mielina (Grapp et al., 2012; Steinfeld et al., 2009). Hemos identificado dos mutaciones nuevas en este gen, una *nonsense* (p.Tyr107Ter) que generaría una proteína truncada que carecería de los dominios de unión al folato, y un cambio previsiblemente de tipo *missense* (p.Arg208Cys) predicho como patogénico por todos los algoritmos utilizados, ya que afecta a un residuo aminoacídico muy conservado entre proteínas ortólogas. Los estudios bioquímicos realizados posteriormente confirmaron la deficiencia en *FOLR1*.

De los cinco pacientes con sospecha de glucogenosis, dos de ellos con afectación hepática presentaron mutaciones en los genes *ALDOB* y *LIPA* y tres con afectación muscular y cardíaca presentaron mutaciones en *CPT2*, *NKX2-5* y *ANO5*.

El P45 presentaba principalmente disfunción hepática por lo que se sospechó de una glucogenosis hepática. Sin embargo, detectamos en homocigosis la mutación más frecuentemente descrita en *ALDOB*. Este gen codifica para la fructosa-1,6-bisfosfato aldolasa B que se expresa en hígado,

intestino delgado y riñón (Salvatore et al., 1986). Mutaciones en este gen producen intolerancia hereditaria a la fructosa (HFI, *hereditary fructose intolerance*) (MIM 612724), una enfermedad autosómica recesiva que debuta con hipoglucemia que puede llegar a ser muy severa, vómitos recurrentes y dolor abdominal similar a la glucogenosis hepática (Santer et al., 2005). La aldolasa B cataliza la transformación de fructosa-1,6-bisfosfato o fructosa-1-fosfato a triosas-fosfato, una etapa central en las rutas de glucólisis y gluconeogénesis (Esposito et al., 2010). La deficiencia en esta enzima produce una depleción de adenosina trifosfato (ATP) y de fosfato inorgánico necesario para la glucógeno fosforilasa hepática, lo que también afecta a la glucogenolisis, provocando una severa hipoglucemia tras la ingesta de fructosa y ausencia de respuesta a la administración de glucagón, características típicas también de la glucogenosis (Van Den Berghe et al., 1973). Cabe resaltar que la deficiencia en la *ALDOB* produce una acumulación de fructosa-1-fosfato, un potente inhibidor competitivo de una enzima de la glicosilación de proteínas, lo que explica el patrón anómalo de las isoformas de la transferrina en estos pacientes, similar al obtenido en los pacientes CDG-I, por lo que a nivel bioquímico también puede sospecharse un CDG (Jaeken et al., 1996). El diagnóstico de la HFI se basa en la respuesta a la sobrecarga de fructosa o en la medida de la actividad enzimática, lo que requiere una biopsia hepática o intestinal (Bouteldja and Timson, 2010; Steinmann and Gitzelmann, 1981). Sin embargo, puesto que ambos métodos son relativamente invasivos y representan un riesgo para el paciente, es preferible verificar la sospecha clínica realizando un diagnóstico genético.

El P48, también con afectación hepática presentó dos mutaciones previamente descritas en el gen *LIPA* (MIM 613497). Este gen codifica para la lipasa lisosomal ácida A (LAL) que cataliza la hidrólisis de ésteres de colesterol y triglicéridos contenidos en las lipoproteínas de baja densidad (LDL). La deficiencia en LAL debuta con hepatomegalia, esteatosis hepática, hipertransaminasemia y elevación de colesterol en sangre, también presentes en las formas hepáticas de glucogenosis. Sin embargo, el perfil atípico de lípidos en sangre que presentó este paciente, sin elevación de colesterol, ni LDL, ni triglicéridos y sin disminución de lipoproteínas de alta densidad (HDL), hizo descartar inicialmente una deficiencia en LAL.

El P47, con afectación muscular presentó una mutación previamente descrita en homocigosis en el gen *CPT2* (MMI 600650). Este gen codifica para la carnitina palmitoiltransferasa 2, enzima implicada en la translocación de ácidos grasos a la matriz mitocondrial para su beta oxidación. Los hallazgos más relevantes de la enfermedad incluyen rabdomiolisis, elevación de creatina kinasa en suero y fallo renal agudo por necrosis tubular, síntomas similares a la glucogenosis muscular (Zutt et al., 2014). Este paciente presentaba niveles normales en plasma de las acilcarnitinas características de la deficiencia en *CPT2*. La mutación identificada (p.Ser113Leu) se ha detectado en el 65% de los alelos patogénicos de todos los pacientes con la forma adulta y más leve de la enfermedad, caracterizada por episodios de rabdomiolisis potenciados tras el ejercicio físico

prolongado (Deschauer et al., 2005). Es posible que p.Ser113Leu genere una proteína con una actividad residual suficiente para que en situación basal la actividad enzimática no se vea comprometida y la deficiencia sólo se manifieste en situaciones de mayor requerimiento energético, como por ejemplo, durante el ejercicio.

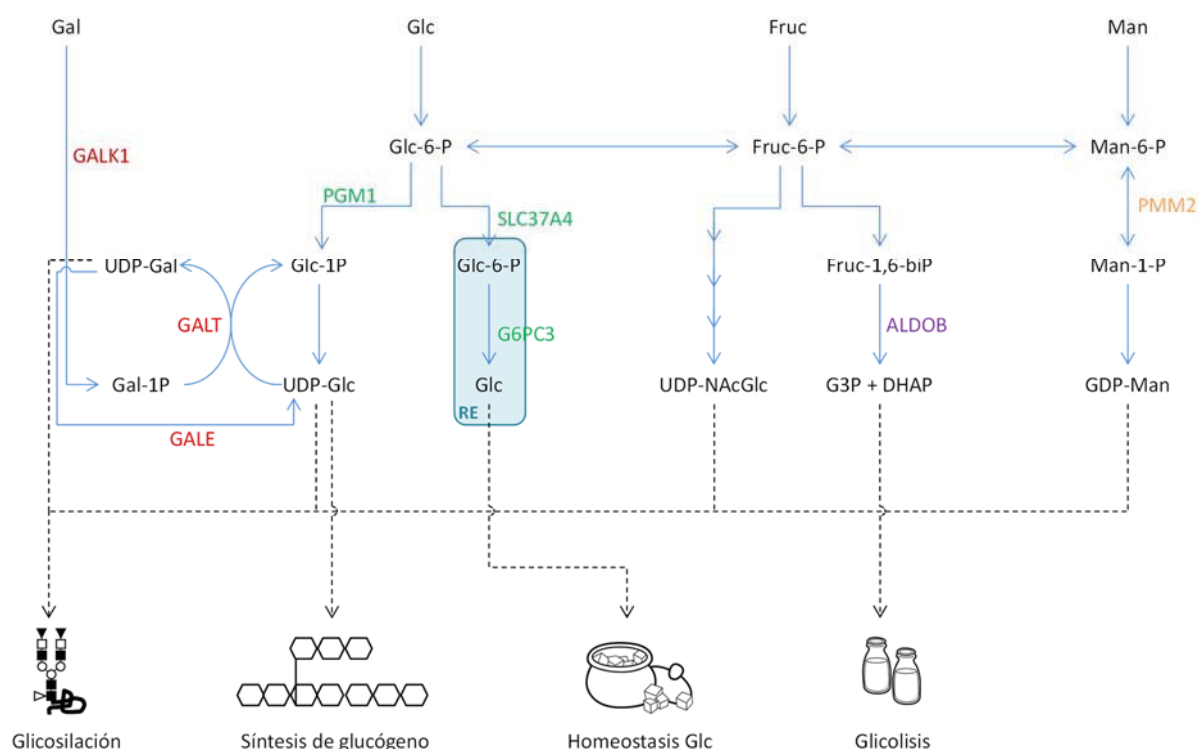
También hemos detectado un paciente (P46) con una sintomatología muscular, que presentaba una mutación en homocigosis en el gen *ANO5* (MIM 608662). Este gen codifica para anoctamina 5, una proteína de función desconocida que se expresa principalmente en músculo cardíaco y esquelético y parece estar implicada en el desarrollo del sistema musculoesquelético (Mizuta et al., 2007). Defectos en este gen producen la distrofia muscular de cintura y extremidades tipo 2L, también llamada anoctaminopatía, una de las formas de distrofia muscular más comunes en adultos (Sarkozy et al., 2013). Se caracteriza principalmente por debilidad muscular de las extremidades proximales y elevación de CK en suero, síntomas característicos también de la glucogenosis muscular. La mutación detectada en el P46 (c.191dupA) representa el 61% de los alelos patogénicos de anoctaminopatía detectados en el norte de Europa (Hicks et al., 2011).

Por último, el P49, también del conjunto de pacientes con sospecha de glucogenosis, presentaba un defecto congénito de corazón. Varios genes asociados a glucogenosis cursan con patología cardíaca exclusiva o en combinación con otros síntomas (Blair et al., 2001). El P49 portaba en homocigosis la mutación previamente descrita p.Arg25Cys en el gen *NKX2-5* (MIM 600584). Hasta la fecha, este paciente es el primero descrito en el que se ha detectado este cambio en homocigosis, lo que podría explicar la patología cardíaca severa y la muerte prematura. *NKX2-5* codifica para un factor de transcripción implicado en la cardiogénesis (Grow and Krieg, 1998; Lyons et al., 1995; Reamon-Buettner and Borlak, 2010) cuya deficiencia produce una cardiopatía severa con herencia autosómica dominante. Se han descrito hasta 70 mutaciones en este gen asociadas a malformaciones congénitas del corazón pero sólo unas pocas se han encontrado en más de una familia no relacionada, entre ellas p.Arg25Cys (Benson et al., 1999; Goldmuntz et al., 2001; McElhinney et al., 2003). Sin embargo, la patogénesis de p.Arg25Cys ha sido cuestionada al haber sido detectada en heterocigosis en población control así como también en pacientes con anomalías cardíacas, por lo que se ha especulado que posea una baja penetrancia (Akcaboy et al., 2008; Alcantara-Ortigoza et al., 2015; De Luca et al., 2011). Es posible que para desarrollar la enfermedad sea necesario que p.Arg25Cys esté en combinación con otro cambio que también produzca un impacto en la proteína.

Estos hallazgos ponen en relieve la dificultad de llegar a un diagnóstico genético aun partiendo de una buena caracterización fenotípica. El amplio espectro fenotípico de los CDG y de la glucogenosis pueden ser solapantes a otras patologías, tal y como ha quedado ilustrado en nuestros resultados.

También debemos tener en cuenta que la deficiencia en una proteína puede tener consecuencias en otras rutas relacionadas. En nuestro caso, ambas patologías pertenecen al metabolismo de carbohidratos. Clásicamente, la ruta metabólica de cada monosacárido se ha considerado de manera independiente y se ha asociado a un grupo de enfermedades aislado: el metabolismo de la galactosa a las galactosemias, el de la glucosa a la glucogenosis, el de la fructosa a intolerancia a la fructosa (HFI) y el de la manosa a la glicosilación de proteínas y lípidos. Sin embargo, todas estas rutas están interrelacionadas (**Figura 24**), lo que justifica que la deficiencia en una de las enzimas pueda tener consecuencias en todas ellas. Este hecho se pone de manifiesto ante algunas de las características fenotípicas solapantes y algunos de los hallazgos bioquímicos comunes entre estos grupos de enfermedades. Este es el caso de las HFI con la glucogenosis y los CDG, tal y como ya se ha discutido, o de las galactosemias con los CDG. Varios autores han descrito deficiencias en algunos genes asociados a la glucogenosis que también tienen alteración en la glicosilación, tales como *PGM1* (MIM 171900), *G6PC3* (MIM 611045), *GAA* (MIM 606800) o *SLC37A4* (MIM 602671) (Hayee et al., 2011; Pérez et al., 2012; Raval et al., 2015). Asimismo, se ha descrito la detección de un paciente PMM2-CDG que clínicamente habían sido diagnosticado como glucogenosis (Choi et al., 2015).

Esta situación pone de manifiesto la necesidad que tienen los laboratorios de diagnóstico de implementar nuevas técnicas de análisis genético que permitan un análisis amplio de genes. De esta manera, ante la identificación de mutaciones en genes implicados en la patología del paciente pero no asociados a la enfermedad a la que venían orientados en un principio, se puede realizar un análisis retrospectivo basándose en los datos fenotípicos del paciente en estrecha colaboración con los médicos para confirmar el diagnóstico genético (*reverse phenotyping*) (Gilissen et al., 2011; Schulze and McMahon, 2004).



**Fig 24. Interrelación del metabolismo de algunos monosacáridos.** En rojo, genes asociados a las galactosemias, en verde a la glucogenosis, en morado a la HFI y en naranja a los CDG.

DHAP, dihidroxiacetona fosfato; Fruc, fructosa; G3P, gliceraldehído-3-fosfato; Gal, galactosa; Glc, glucosa; Man, manosa.

En resumen, un diagnóstico rápido y preciso es importante para poder ofrecer un consejo genético adecuado, un diagnóstico prenatal y/o preimplantacional si es requerido e iniciar un tratamiento lo antes posible, con el fin de evitar daños irreversibles (Katsanis and Katsanis, 2013). Algunos de los diagnósticos genéticos establecidos en este trabajo han permitido la prescripción de un tratamiento específico para su defecto en algunos de los pacientes. Por ejemplo, tras el diagnóstico de P45 se ha procedido al inicio de una dieta libre en fructosa (Berni Canani et al., 2016), el P44 se sometió un tratamiento con ácido fólico (Perez-Duenas et al., 2011), el P47 una dieta restringida en ácidos grasos de cadena larga y suplementada con ácidos grasos de cadena media y uno de los pacientes deficientes en *GALT* comenzó con una dieta libre en galactosa (Timson, 2016). Los otros dos pacientes deficientes en *GALT* habían fallecido antes del diagnóstico. Más llamativa fue la situación del P48, que estaba siendo valorado para trasplante hepático y tras la identificación del defecto genético en *LIPA*, ha sido incluido en un ensayo clínico de reemplazamiento enzimático en fase 3 (Balwani et al., 2013; Valayannopoulos et al., 2014).

## CONCLUSIONES

- 1 La implementación de la secuenciación masiva como método de diagnóstico genético ha resultado eficaz en el diagnóstico de enfermedades genéticamente heterogéneas como los CDG y la glucogenosis, ya que permite analizar muchos genes de manera simultánea. Además, NGS ha resultado ser más sensible que la secuenciación convencional Sanger al evitar el *allele dropout* en dos muestras y permitir la detección de una variación en el número de copia. Además, cabe destacar la sensibilidad de la técnica en la detección de mutaciones identificadas en genes con gran homología de secuencia a otras regiones del genoma.
- 2 Los resultados sugieren que los paneles personalizados de genes candidatos, a pesar de tener una cobertura suficiente para su aplicación al diagnóstico genético, no son adecuados para el estudio de enfermedades emergentes como los CDG donde nuevos genes son descritos con regularidad.
- 3 Los resultados sugieren que el exoma clínico puede ser una aproximación más adecuada para el diagnóstico genético de enfermedades que debutan con una clínica inespecífica y/o carecen de un biomarcador patognomónico, ya que permite un análisis de genes no sesgado y posee una cobertura uniforme y con validez diagnóstica.
- 4 El éxito del diagnóstico está basado en el establecimiento en cada caso de un adecuado filtrado y priorizado de las variantes detectadas durante el análisis de datos de secuenciación masiva y debe estar basado en el conocimiento de la patología.
- 5 Mutaciones patogénicas en el gen *CCDC115* parecen producir un defecto en la glicosilación. Los estudios funcionales realizados sugieren que la proteína que codifica este gen podría estar implicada en la homeostasis del aparato de Golgi y su deficiencia afectaría al correcto funcionamiento del orgánulo y, con ello, a las rutas de N- y O-glicosilación de proteínas.
- 6 Nuestros resultados demuestran el potencial de la secuenciación masiva aplicada al diagnóstico de enfermedades genéticamente heterogéneas permitiendo, no sólo el correcto asesoramiento genético, sino también la aplicación de tratamientos específicos así como dirigir la investigación hacia terapias basadas en los mecanismos de acción de las mutaciones que se identifican (medicina de precisión).



## BIBLIOGRAFÍA

**A**dzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods* 7, 248-249.

Akcaboy, M.I., Cengiz, F.B., Inceoglu, B., Ucar, T., Atalay, S., Tutar, E., and Tekin, M. (2008). The effect of p.Arg25Cys alteration in NKX2-5 on conotruncal heart anomalies: mutation or polymorphism? *Pediatr Cardiol* 29, 126-129.

Alcantara-Ortigoza, M.A., De Rubens-Figueroa, J., Reyna-Fabian, M.E., Estandia-Ortega, B., Gonzalez-del Angel, A., Molina-Alvarez, B., Velazquez-Aragon, J.A., Villagomez-Martinez, S., Pereira-Lopez, G.I., Martinez-Cruz, V., *et al.* (2015). Germline mutations in NKX2-5, GATA4, and CRELD1 are rare in a Mexican sample of Down syndrome patients with endocardial cushion and septal heart defects. *Pediatr Cardiol* 36, 802-808.

Ambardar, S., Gupta, R., Trakroo, D., Lal, R., and Vakhlu, J. (2016). High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian J Microbiol* 56, 394-404.

Ashida, H., Maeda, Y., and Kinoshita, T. (2006). DPM1, the catalytic subunit of dolichol-phosphate mannose synthase, is tethered to and stabilized on the endoplasmic reticulum membrane by DPM3. *The Journal of biological chemistry* 281, 896-904.

**B**ahassi el, M., and Stambrook, P.J. (2014). Next-generation sequencing technologies: breaking the sound barrier of human genetics. *Mutagenesis* 29, 303-310.

Balwani, M., Breen, C., Enns, G.M., Deegan, P.B., Honzik, T., Jones, S., Kane, J.P., Malinova, V., Sharma, R., Stock, E.O., *et al.* (2013). Clinical effect and safety profile of recombinant human lysosomal acid lipase in patients with cholesteryl ester storage disease. *Hepatology* 58, 950-957.

Basiri, K., Belaya, K., Liu, W., Maxwell, S., Sedghi, M., and Beeson, D. (2013). Clinical features in a large Iranian family with a limb-girdle congenital myasthenic syndrome due to a mutation in DPAGT1. *Neuromuscular Disorders* 23, 469-472.

Bean, P., Sutphin, M.S., Liu, Y., Anton, R., Reynolds, T.B., Shoenfeld, Y., and Peter, J.B. (1995). Carbohydrate-deficient transferrin and false-positive results for alcohol abuse in primary biliary cirrhosis: differential diagnosis by detection of mitochondrial autoantibodies. *Clin Chem* 41, 858-861.

Belaya, K., Finlayson, S., Slater, C.R., Cossins, J., Liu, W., Maxwell, S., McGowan, S.J., Maslau, S., Twigg, S., Walls, T.J., *et al.* (2012). Mutations in DPAGT1 Cause a Limb-Girdle Congenital Myasthenic Syndrome with Tubular Aggregates. *The American Journal of Human Genetics* 91, 193-201.

Benson, D.W., Silberbach, G.M., Kavanaugh-McHugh, A., Cottrill, C., Zhang, Y., Riggs, S., Smalls, O., Johnson, M.C., Watson, M.S., Seidman, J.G., *et al.* (1999). Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. *J Clin Invest* 104, 1567-1573.

Berni Canani, R., Pezzella, V., Amoroso, A., Cozzolino, T., Di Scala, C., and Passariello, A. (2016). Diagnosing and Treating Intolerance to Carbohydrates in Children. *Nutrients* 8, 157.

- Bexiga, M.G., and Simpson, J.C. (2013). Human diseases associated with form and function of the Golgi complex. *Int J Mol Sci* **14**, 18670-18681.
- Bissar-Tadmouri, N., Donahue, W.L., Al-Gazali, L., Nelson, S.F., Bayrak-Toydemir, P., and Kantarci, S. (2014). X chromosome exome sequencing reveals a novel ALG13 mutation in a nonsyndromic intellectual disability family with multiple affected male siblings. *American Journal of Medical Genetics Part A* **164**, 164-169.
- Blair, E., Redwood, C., Ashrafian, H., Oliveira, M., Broxholme, J., Kerr, B., Salmon, A., Ostman-Smith, I., and Watkins, H. (2001). Mutations in the gamma(2) subunit of AMP-activated protein kinase cause familial hypertrophic cardiomyopathy: evidence for the central role of energy compromise in disease pathogenesis. *Hum Mol Genet* **10**, 1215-1220.
- Blais, J., Lavoie, S.B., Giroux, S., Bussieres, J., Lindsay, C., Dionne, J., Laroche, M., Giguere, Y., and Rousseau, F. (2015). Risk of Misdiagnosis Due to Allele Dropout and False-Positive PCR Artifacts in Molecular Diagnostics: Analysis of 30,769 Genotypes. *J Mol Diagn* **17**, 505-514.
- Bouteldja, N., and Timson, D.J. (2010). The biochemical basis of hereditary fructose intolerance. *J Inherit Metab Dis* **33**, 105-112.
- Buczowska, A., Swiezewska, E., and Lefeber, D.J. (2014). Genetic defects in dolichol metabolism. *Journal of Inherited Metabolic Disease* **38**, 157-169.
- Carss, K.J., Stevens, E., Foley, R.A., Cirak, S., Riemersma, M., Torelli, S., Hoischen, A., Willer, T., van Scherpenzeel, M., Moore, S.A., *et al.* (2013). Mutations in GDP-Mannose Pyrophosphorylase B Cause Congenital and Limb-Girdle Muscular Dystrophies Associated with Hypoglycosylation of  $\alpha$ -Dystroglycan. *The American Journal of Human Genetics* **93**, 29-41.
- Claes, K.B., and De Leeneer, K. (2014). Dealing with pseudogenes in molecular diagnostics in the next-generation sequencing era. *Methods in Molecular Biology* **1167**, 303-315.
- Codina-Sola, M., Rodriguez-Santiago, B., Homs, A., Santoyo, J., Rigau, M., Aznar-Lain, G., Del Campo, M., Gener, B., Gabau, E., Botella, M.P., *et al.* (2015). Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum disorders. *Mol Autism* **6**, 21.
- Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65.
- Cossins, J., Belaya, K., Hicks, D., Salih, M.A., Finlayson, S., Carboni, N., Liu, W., Maxwell, S., Zoltowska, K., Farsani, G., *et al.* (2013). Congenital myasthenic syndromes due to mutations in ALG2 and ALG14. *Brain* **136**, 944-956.
- Charlwood, J., Clayton, P., Keir, G., Mian, N., and Winchester, B. (1998). Defective galactosylation of serum transferrin in galactosemia. *Glycobiology* **8**, 351-357.

- Chen, Y.T., Bazzarre, C.H., Lee, M.M., Sidbury, J.B., and Coleman, R.A. (1993). Type I glycogen storage disease: nine years of management with cornstarch. *Eur J Pediatr* 152 Suppl 1, S56-59.
- Cheng, A., Zhang, M., Gentry, M.S., Worby, C.A., Dixon, J.E., and Saltiel, A.R. (2007). A role for AGL ubiquitination in the glycogen storage disorders of Lafora and Cori's disease. *Genes Dev* 21, 2399-2409.
- Choi, M., Scholl, U.I., Ji, W., Liu, T., Tikhonova, I.R., Zumbo, P., Nayir, A., Bakkaloğlu, A., Ozen, S., Sanjad, S., *et al.* (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 106, 19096-19101.
- Choi, R., Woo, H., Choe, B.-H., Park, S., Yoon, Y., Ki, C.-S., Lee, S.-Y., Kim, J.-W., Song, J., Kim, D., *et al.* (2015). Application of whole exome sequencing to a rare inherited metabolic disease with neurological and gastrointestinal manifestations: A congenital disorder of glycosylation mimicking glycogen storage disease. *Clinica Chimica Acta* 444, 50-53.
- D**avit-Spraul, A., Piraud, M., Dobbelaere, D., Valayannopoulos, V., Labrune, P., Habes, D., Bernard, O., Jacquemin, E., and Baussan, C. (2011). Liver glycogen storage diseases due to phosphorylase system deficiencies: diagnosis thanks to non invasive blood enzymatic and molecular studies. *Mol Genet Metab* 104, 137-143.
- de Loos, F., Huijben, K.M., van der Kar, N.C., Monnens, L.A., van den Heuvel, L.P., Groener, J.E., de Moor, R.A., and Wevers, R.A. (2002). Hemolytic uremic syndrome attributable to *Streptococcus pneumoniae* infection: a novel cause for secondary protein N-glycan abnormalities. *Clin Chem* 48, 781-784.
- De Luca, A., Sarkozy, A., Ferese, R., Consoli, F., Lepri, F., Dentici, M.L., Vergara, P., De Zorzi, A., Versacci, P., Digilio, M.C., *et al.* (2011). New mutations in ZFPM2/FOG2 gene in tetralogy of Fallot and double outlet right ventricle. *Clin Genet* 80, 184-190.
- Denecke, J., and Kranz, C. (2009). Hypoglycosylation due to dolichol metabolism defects. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1792, 888-895.
- Deschauer, M., Wieser, T., and Zierz, S. (2005). Muscle carnitine palmitoyltransferase II deficiency: clinical and molecular genetic features and diagnostic aspects. *Arch Neurol* 62, 37-41.
- Dörre, K., Olczak, M., Wada, Y., Sosicka, P., Grüneberg, M., Reunert, J., Kurlemann, G., Fiedler, B., Biskup, S., Hörtnagel, K., *et al.* (2015). A new case of UDP-galactose transporter deficiency (SLC35A2-CDG): molecular basis, clinical phenotype, and therapeutic approach. *Journal of Inherited Metabolic Disease* 38, 931-940.
- Dorschner, M.O., Amendola, L.M., Turner, E.H., Robertson, P.D., Shirts, B.H., Gallego, C.J., Bennett, R.L., Jones, K.L., Tokita, M.J., Bennett, J.T., *et al.* (2013). Actionable, Pathogenic Incidental Findings in 1,000 Participants' Exomes. *The American Journal of Human Genetics* 93, 631-640.

- E**sposito, G., Imperato, M.R., Ieno, L., Sorvillo, R., Benigno, V., Parenti, G., Parini, R., Vitagliano, L., Zagari, A., and Salvatore, F. (2010). Hereditary fructose intolerance: functional study of two novel ALDOB natural variants and characterization of a partial gene deletion. *Hum Mutat* 31, 1294-1303.
- Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8, 175-185.
- F**eng, Y., Chen, D., Wang, G.L., Zhang, V.W., and Wong, L.J. (2015). Improved molecular diagnosis by the detection of exonic deletions with target gene capture and deep sequencing. *Genet Med* 17, 99-107.
- Forgac, M. (2007). Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. *Nat Rev Mol Cell Biol* 8, 917-929.
- Foulquier, F. (2008). COG defects, birth and rise! *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1792, 896-902.
- Foulquier, F., Amyere, M., Jaeken, J., Zeevaert, R., Schollen, E., Race, V., Bammens, R., Morelle, W., Rosnoblet, C., Legrand, D., *et al.* (2012). TMEM165 deficiency causes a congenital disorder of glycosylation. *American journal of human genetics* 91, 15-26.
- Freeze, H.H. (2006). Genetic defects in the human glycome. *Nature Reviews Genetics* 7, 537-551.
- Freeze, H.H., Chong, J.X., Bamshad, M.J., and Ng, B.G. (2014). Solving Glycosylation Disorders: Fundamental Approaches Reveal Complicated Pathways. *The American Journal of Human Genetics* 94, 161-175.
- Freeze, H.H., and Elbein, A.D. (2009). Glycosylation Precursors. In *Essentials of Glycobiology*, A. Varki, R.D. Cummings, J.D. Esko, H.H. Freeze, P. Stanley, C.R. Bertozzi, G.W. Hart, and M.E. Etzler, eds. (Cold Spring Harbor (NY)).
- G**arcía-Silva, M.T., Matthijs, G., Schollen, E., Cabrera, J.C., Sanchez del Pozo, J., Martí Herreros, M., Simón, R., Maties, M., Martín Hernández, E., Hennet, T., *et al.* (2004). Congenital disorder of glycosylation (CDG) type 1e. A new patient. *J Inherit Metab Dis* 27, 591-600.
- Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., *et al.* (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature* 511, 344-347.
- Gilissen, C., Hoischen, A., Brunner, H.G., and Veltman, J.A. (2011). Unlocking Mendelian disease using exome sequencing. *Genome Biol* 12, 228.
- Gilissen, C., Hoischen, A., Brunner, H.G., and Veltman, J.A. (2012). Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* 20, 490-497.
- Goldmuntz, E., Geiger, E., and Benson, D.W. (2001). NKX2.5 mutations in patients with tetralogy of fallot. *Circulation* 104, 2565-2568.

- Goldstein, J.L., Austin, S.L., Boyette, K., Kanaly, A., Veerapandiyan, A., Rehder, C., Kishnani, P.S., and Bali, D.S. (2010). Molecular analysis of the AGL gene: identification of 25 novel mutations and evidence of genetic heterogeneity in patients with Glycogen Storage Disease Type III. *Genet Med* 12, 424-430.
- Grapp, M., Just, I.A., Linnankivi, T., Wolf, P., Lucke, T., Hausler, M., Gartner, J., and Steinfeld, R. (2012). Molecular characterization of folate receptor 1 mutations delineates cerebral folate transport deficiency. *Brain* 135, 2022-2031.
- Grow, M.W., and Krieg, P.A. (1998). Tinman function is essential for vertebrate heart development: elimination of cardiac differentiation by dominant inhibitory mutants of the tinman-related genes, *XNkx2-3* and *XNkx2-5*. *Dev Biol* 204, 187-196.
- Guillard, M., Gloerich, J., Wessels, H., Morava, E., Wevers, R.A., and Lefeber, D.J. (2009). Automated measurement of permethylated serum N-glycans by MALDI-linear ion trap mass spectrometry. *Carbohydrate Research* 344, 1550-1557.
- Guillard, M., Morava, E., van Delft, F.L., Hague, R., Körner, C., Adamowicz, M., Wevers, R.A., and Lefeber, D.J. (2011a). Plasma N-Glycan Profiling by Mass Spectrometry for Congenital Disorders of Glycosylation Type II. *Clinical Chemistry* 57, 593-602.
- Guillard, M., Wada, Y., Hansikova, H., Yuasa, I., Vesela, K., Ondruskova, N., Kadoya, M., Janssen, A., den Heuvel, L., Morava, E., *et al.* (2011b). Transferrin mutations at the glycosylation site complicate diagnosis of congenital disorders of glycosylation type I. *Journal of Inherited Metabolic Disease* 34, 901-906.
- Hackler, R., Arndt, T., Kleine, T.O., and Gressner, A.M. (1995). Effect of separation conditions on automated isoelectric focusing of carbohydrate-deficient transferrin and other human isotransferrins using the PhastSystem. *Analytical biochemistry* 230, 281-289.
- Haeuptle, M.A., Hülsmeier, A.J., and Hennet, T. (2010). HPLC and mass spectrometry analysis of dolichol-phosphates at the cell culture scale. *Analytical Biochemistry* 396, 133-138.
- Hayee, B.H., Antonopoulos, A., Murphy, E.J., Rahman, F.Z., Sewell, G., Smith, B.N., McCartney, S., Furman, M., Hall, G., Bloom, S.L., *et al.* (2011). G6PC3 mutations are associated with a major defect of glycosylation: a novel mechanism for neutrophil dysfunction. *Glycobiology* 21, 914-924.
- Hayes, B.K., and Varki, A. (1993). The biosynthesis of oligosaccharides in intact Golgi preparations from rat liver. Analysis of N-linked and O-linked glycans labeled by UDP-[6-3H]N-acetylgalactosamine. *The Journal of biological chemistry* 268, 16170-16178.
- Hays, A.P., Hallett, M., Delfs, J., Morris, J., Sotrel, A., Shevchuk, M.M., and DiMauro, S. (1981). Muscle phosphofructokinase deficiency: abnormal polysaccharide in a case of late-onset myopathy. *Neurology* 31, 1077-1086.
- Helander, A., Wielders, J.P.M., Jeppsson, J.-O., Weykamp, C., Siebelder, C., Anton, R.F., Schellenberg, F., Whitfield, J.B., and *on of, I.* (2010). Toward standardization of carbohydrate-deficient transferrin (CDT) measurements: II. Performance of a laboratory network running the HPLC candidate reference measurement procedure and evaluation of a candidate reference material.

- Clinical Chemistry and Laboratory Medicine *48*, 1585-1592.
- Heller, S., Worona, L., and Consuelo, A. (2008). Nutritional therapy for glycogen storage diseases. *J Pediatr Gastroenterol Nutr* *47 Suppl 1*, S15-21.
- Hennet, T. (2012). Diseases of glycosylation beyond classical congenital disorders of glycosylation. *Biochimica et Biophysica Acta (BBA) - General Subjects* *1820*, 1306-1317.
- Hennet, T., and Cabalzar, J. (2015). Congenital disorders of glycosylation: a concise chart of glycoalyx dysfunction. *Trends in Biochemical Sciences* *40*, 377-384.
- Hicks, D., Sarkozy, A., Muelas, N., Koehler, K., Huebner, A., Hudson, G., Chinnery, P.F., Barresi, R., Eagle, M., Polvikoski, T., *et al.* (2011). A founder mutation in Anoctamin 5 is a major cause of limb-girdle muscular dystrophy. *Brain* *134*, 171-182.
- Imtiaz, F., Al-Mostafa, A., and Al-Hassnan, Z. (2012). Further Delineation of the Phenotype of Congenital Disorder of Glycosylation DPAGT1-CDG (CDG-Ij) Identified by Homozygosity Mapping. *JIMD Reports* *2*, 107-111.
- Iqbal, Z., Shahzad, M., Vissers, L., van Scherpenzeel, M., Gilissen, C., Razzaq, A., Zahoor, M., Khan, S.N., Kleefstra, T., Veltman, J.A., *et al.* (2012). A compound heterozygous mutation in DPAGT1 results in a congenital disorder of glycosylation with a relatively mild phenotype. *European Journal of Human Genetics* *21*, 844-849.
- Jaeken, J. (2010). Congenital disorders of glycosylation. *Annals of the New York Academy of Sciences* *1214*, 190-198.
- Jaeken, J., Hennet, T., Matthijs, G., and Freeze, H.H. (2009). CDG nomenclature: Time for a change! *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* *1792*, 825-826.
- Jaeken, J., Pirard, M., Adamowicz, M., Pronicka, E., and van Schaftingen, E. (1996). Inhibition of phosphomannose isomerase by fructose 1-phosphate: an explanation for defective N-glycosylation in hereditary fructose intolerance. *Pediatr Res* *40*, 764-766.
- Jansen, E.J., Timal, S., Ryan, M., Ashikov, A., van Scherpenzeel, M., Graham, L.A., Mandel, H., Hoischen, A., Iancu, T.C., Raymond, K., *et al.* (2016a). ATP6AP1 deficiency causes an immunodeficiency with hepatopathy, cognitive impairment and abnormal protein glycosylation. *Nat Commun* *7*, 11600.
- Jansen, J.C., Cirak, S., van Scherpenzeel, M., Timal, S., Reunert, J., Rust, S., Perez, B., Vicogne, D., Krawitz, P., Wada, Y., *et al.* (2016b). CCDC115 Deficiency Causes a Disorder of Golgi Homeostasis with Abnormal Protein Glycosylation. *Am J Hum Genet* *98*, 310-321.
- Jansen, J.C., Timal, S., van Scherpenzeel, M., Michelakakis, H., Vicogne, D., Ashikov, A., Moraitou, M., Hoischen, A., Huijben, K., Steenbergen, G., *et al.* (2016c). TMEM199 Deficiency Is a Disorder of Golgi Homeostasis Characterized by Elevated Aminotransferases, Alkaline Phosphatase, and Cholesterol and Abnormal Glycosylation. *Am J Hum Genet* *98*, 322-330.
- Jarinova, O., and Ekker, M. (2012). Regulatory variations in the era of next-generation sequencing: Implications for clinical molecular diagnostics. *Human Mutation* *33*, 1021-1030.



- Jones, M.A., Bhide, S., Chin, E., Ng, B.G., Rhodenizer, D., Zhang, V.W., Sun, J.J., Tanner, A., Freeze, H.H., and Hegde, M.R. (2011). Targeted polymerase chain reaction-based enrichment and next generation sequencing for diagnostic testing of congenital disorders of glycosylation. *Genetics in Medicine* 13, 921-932.
- Jones, M.A., Rhodenizer, D., da Silva, C., Huff, I.J., Keong, L., Bean, L.J.H., Coffee, B., Collins, C., Tanner, A.K., He, M., *et al.* (2013). Molecular diagnostic testing for congenital disorders of glycosylation (CDG): detection rate for single gene testing and next generation sequencing panel testing. *Molecular genetics and metabolism* 110, 78-85.
- K**atsanis, S.H., and Katsanis, N. (2013). Molecular genetic testing and the future of clinical genomics. *Nature reviews Genetics* 14, 415-426.
- Kishnani, P.S., Nicolino, M., Voit, T., Rogers, R.C., Tsai, A.C., Waterson, J., Herman, G.E., Amalfitano, A., Thurberg, B.L., Richards, S., *et al.* (2006). Chinese hamster ovary cell-derived recombinant human acid alpha-glucosidase in infantile-onset Pompe disease. *J Pediatr* 149, 89-97.
- Koboldt, D.C., Ding, L., Mardis, E.R., and Wilson, R.K. (2010). Challenges of sequencing human genomes. *Briefings in Bioinformatics* 11, 484-498.
- Koboldt, D.C., Steinberg, K., Larson, D.E., Wilson, R.K., and Mardis, E.R. (2013). The next-generation sequencing revolution and its impact on genomics. *Cell* 155, 27-38.
- Koehler, K., Malik, M., Mahmood, S., Gießelmann, S., Beetz, C., Hennings, J.C., Huebner, A.K., Grahn, A., Reunert, J., Nürnberg, G., *et al.* (2013). Mutations in GMPPA cause a glycosylation disorder characterized by intellectual disability and autonomic dysfunction. *American journal of human genetics* 93, 727-734.
- Körner, C., Lehle, L., and von Figura, K. (1998). Abnormal synthesis of mannose 1-phosphate derived carbohydrates in carbohydrate-deficient glycoprotein syndrome type I fibroblasts with phosphomannomutase deficiency. *Glycobiology* 8, 165-171.
- Kreuder, J., Borkhardt, A., Repp, R., Pekrun, A., Gottsche, B., Gottschalk, U., Reichmann, H., Schachenmayr, W., Schlegel, K., and Lampert, F. (1996). Brief report: inherited metabolic myopathy and hemolysis due to a mutation in aldolase A. *N Engl J Med* 334, 1100-1104.
- Kuroda, Y., Hamaguchi, R., Moriyama, K., Tanimoto, T., and Haginaka, J. (2013). Improved capillary electrophoresis method for the analysis of carbohydrate-deficient transferrin in human serum, avoiding interference by complement C3. *Journal of Pharmaceutical and Biomedical Analysis* 76, 81-86.
- L**efeber, D.J., Morava, E., and Jaeken, J. (2011). How to find and diagnose a CDG due to defective N-glycosylation. *Journal of Inherited Metabolic Disease* 34, 849-852.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.



- Lohi, H., Ianzano, L., Zhao, X.C., Chan, E.M., Turnbull, J., Scherer, S.W., Ackerley, C.A., and Minassian, B.A. (2005). Novel glycogen synthase kinase 3 and ubiquitination pathways in progressive myoclonus epilepsy. *Hum Mol Genet* *14*, 2727-2736.
- Loman, N.J., Constantinidou, C., Chan, J.Z., Halachev, M., Sergeant, M., Penn, C.W., Robinson, E.R., and Pallen, M.J. (2012). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol* *10*, 599-606.
- Losfeld, M., Ng, B.G., Kircher, M., Buckingham, K.J., Turner, E.H., Eroshkin, A., Smith, J.D., Shendure, J., Nickerson, D.A., Bamshad, M.J., *et al.* (2014). A new congenital disorder of glycosylation caused by a mutation in SSR4, the signal sequence receptor 4 protein of the TRAP complex. *Human Molecular Genetics* *23*, 1602-1605.
- Lucchiari, S., Pagliarani, S., Salani, S., Filocamo, M., Di Rocco, M., Melis, D., Rodolico, C., Musumeci, O., Toscano, A., Bresolin, N., *et al.* (2006). Hepatic and neuromuscular forms of glycogenosis type III: nine mutations in AGL. *Hum Mutat* *27*, 600-601.
- Lyons, I., Parsons, L.M., Hartley, L., Li, R., Andrews, J.E., Robb, L., and Harvey, R.P. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes Dev* *9*, 1654-1666.
- Maeda, Y., and Kinoshita, T. (2010). The Acidic Environment of the Golgi Is Critical for Glycosylation and Transport. *Methods in Enzymology* *480*, 495-510.
- Maeda, Y., Tomita, S., Watanabe, R., Ohishi, K., and Kinoshita, T. (1998). DPM2 regulates biosynthesis of dolichol phosphate-mannose in mammalian cells: correct subcellular localization and stabilization of DPM1, and binding of dolichol phosphate. *The EMBO journal* *17*, 4920-4929.
- Mardis, E.R. (2013). Next-Generation Sequencing Platforms. *Annual review of analytical chemistry (Palo Alto, Calif)* *6*, 287-303.
- Marquardt, T., Ullrich, K., Zimmer, P., Hasilik, A., Deufel, T., and Harms, E. (1995). Carbohydrate-deficient glycoprotein syndrome (CDGS)--glycosylation, folding and intracellular transport of newly synthesized glycoproteins. *Eur J Cell Biol* *66*, 268-273.
- Marsh, M., Tu, O., Dolnik, V., Roach, D., Solomon, N., Bechtol, K., Smietana, P., Wang, L., Li, X., Cartwright, P., *et al.* (1997). High-throughput DNA sequencing on a capillary array electrophoresis system. *J Capillary Electrophor* *4*, 83-89.
- Martinez-Duncker, I., Dupré, T., Piller, V., Candeliere, J.J., Trichet, C., Tchernia, G., Oriol, R., and Mollicone, R. (2005). Genetic complementation reveals a novel human congenital disorder of glycosylation of type II, due to inactivation of the Golgi CMP-sialic acid transporter. *Blood* *105*, 2671-2676.

**M**aeda, Y., and Kinoshita, T. (2008). Dolichol-phosphate mannose synthase: Structure, function and regulation. *Biochimica et Biophysica Acta (BBA) - General Subjects* *1780*, 861-868.

- Matalonga, L., Bravo, M., Peinado, C.S., Garcia-Pelegri, E., Ugarteburu, O., Vidal, S., Llambrich, M., Quintana, E., Fuster-Jorge, P., Gonzalez-Bravo, M.N., *et al.* (2016). Mutations in TRAPPC11 are Associated with a Congenital Disorder of Glycosylation. *Hum Mutat*, In press.
- Matthijs, G., Rymen, D., Millón, M., Souche, E., and Race, V. (2012). Approaches to homozygosity mapping and exome sequencing for the identification of novel types of CDG. *Glycoconjugate Journal* 30, 67-76.
- McArdle, B., and Verel, D. (1956). Responses to ischaemic work in the human forearm. *Clin Sci (Lond)* 15, 305-318.
- McElhinney, D.B., Geiger, E., Blinder, J., Benson, D.W., and Goldmuntz, E. (2003). NKX2.5 mutations in patients with congenital heart disease. *J Am Coll Cardiol* 42, 1650-1655.
- Metzker, M.L. (2009). Sequencing technologies — the next generation. *Nature Reviews Genetics* 11, 31-46.
- Mills, P.B., Mills, K., Mian, N., Winchester, B.G., and Clayton, P.T. (2003). Mass spectrometric analysis of glycans in elucidating the pathogenesis of CDG type IIx. *Journal of Inherited Metabolic Disease* 26, 119-134.
- Mizuta, K., Tsutsumi, S., Inoue, H., Sakamoto, Y., Miyatake, K., Miyawaki, K., Noji, S., Kamata, N., and Itakura, M. (2007). Molecular characterization of GDD1/TMEM16E, the gene product responsible for autosomal dominant gnathodiaphyseal dysplasia. *Biochem Biophys Res Commun* 357, 126-132.
- Mohamed, M., Cantagrel, V., Al-Gazali, L., Wevers, R.A., Lefeber, D.J., and Morava, E. (2011). Normal glycosylation screening does not rule out SRD5A3-CDG. *European Journal of Human Genetics* 19, 1019.
- Morava, E. (2014). Galactose supplementation in phosphoglucomutase-1 deficiency; review and outlook for a novel treatable CDG. *Molecular genetics and metabolism* 112, 275-279.
- Morava, É., Guillard, M., Lefeber, D.J., and Wevers, R.A. (2009). Autosomal recessive cutis laxa syndrome revisited. *European Journal of Human Genetics* 17, 1099-1110.
- Morava, E., Rahman, S., Peters, V., Baumgartner, M.R., Patterson, M., and Zschocke, J. (2015). Quo vadis: the re-definition of "inborn metabolic diseases". *J Inherit Metab Dis* 38, 1003-1006.
- Morava, E., Vodopiutz, J., Lefeber, D.J., Janecke, A.R., Schmidt, W.M., Lechner, S., Item, C.B., Sykut-Cegielska, J., Adamowicz, M., Wierzba, J., *et al.* (2012). Defining the Phenotype in Congenital Disorder of Glycosylation Due to ALG1 Mutations. *Pediatrics* 130, e1034-1039.
- Morava, E., Zeevaert, R., Korsch, E., Huijben, K., Wopereis, S., Matthijs, G., Keymolen, K., Lefeber, D.J., Meirleir, L., and Wevers, R.A. (2007). A common mutation in the COG7 gene with a consistent phenotype including microcephaly, adducted thumbs, growth retardation, VSD and episodes of hyperthermia. *European Journal of Human Genetics* 15, 638-645.
- Morelle, W., and Michalski, J.-C. (2007). Analysis of protein glycosylation by mass spectrometry. *Nature Protocols* 2, 1585-1602.
- Mullis, K.B., and Faloona, F.A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in enzymology* 155, 335-350.

Munsat, T.L. (1970). A standardized forearm ischemic exercise test. *Neurology* 20, 1171-1178.

Murawaki, Y., Sugisaki, H., Yuasa, I., and Kawasaki, H. (1997). Serum carbohydrate-deficient transferrin in patients with nonalcoholic liver disease and with hepatocellular carcinoma. *Clin Chim Acta* 259, 97-108.

**N**g, B.G., Buckingham, K.J., Raymond, K., Kircher, M., Turner, E.H., He, M., Smith, J.D., Eroshkin, A., Szybowska, M., Losfeld, M.E., *et al.* (2013). Mosaicism of the UDP-Galactose Transporter SLC35A2 Causes a Congenital Disorder of Glycosylation. *The American Journal of Human Genetics* 92, 632-636.

Ng, B.G., Kranz, C., Hagebeuk, E.E.O., Duran, M., Abeling, N., Wuyts, B., Ungar, D., Lupashin, V., Hartdorff, C.M., Poll-The, B.T., *et al.* (2007). Molecular and clinical characterization of a Moroccan Cog7 deficient patient. *Molecular genetics and metabolism* 91, 201-204.

Ng, P.C., and Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Res* 11, 863-874.

Niehues, R., Hasilik, M., Alton, G., Korner, C., Schiebe-Sukumar, M., Koch, H.G., Zimmer, K.P., Wu, R., Harms, E., Reiter, K., *et al.* (1998). Carbohydrate-deficient glycoprotein syndrome type Ib. Phosphomannose isomerase deficiency and mannose therapy. *J Clin Invest* 101, 1414-1420.

Nord, A.S., Lee, M., King, M.C., and Walsh, T. (2011). Accurate and exact CNV identification from targeted high-throughput sequence data. *BMC Genomics* 12, 184.

**O**'Donnell-Luria, A.H., and Miller, D.T. (2016). A Clinician's perspective on clinical exome sequencing. *Hum Genet* 135, 643-654.

Okazaki, T., Murata, M., Kai, M., Adachi, K., Nakagawa, N., Kasagi, N., Matsumura, W., Maegaki, Y., and Nanba, E. (2016). Clinical Diagnosis of Mendelian Disorders Using a Comprehensive Gene-Targeted Panel Test for Next-Generation Sequencing. *Yonago Acta Med* 59, 118-125.

Okubo, M., Kanda, F., Horinishi, A., Takahashi, K., Okuda, S., Chihara, K., and Murase, T. (1999). Glycogen storage disease type IIIa: first report of a causative missense mutation (G1448R) of the glycogen debranching enzyme gene found in a homozygous patient. *Hum Mutat* 14, 542-543.

Ozen, H. (2007). Glycogen storage diseases: new perspectives. *World J Gastroenterol* 13, 2541-2553.

**P**agani, F., Pariyarath, R., Garcia, R., Stuani, C., Burlina, A.B., Ruotolo, G., Rabusin, M., and Baralle, F.E. (1998). New lysosomal acid lipase gene mutants explain the phenotype of Wolman disease and cholesteryl ester storage disease. *J Lipid Res* 39, 1382-1388.

Panneerselvam, K., and Freeze, H.H. (1995). Enzymes involved in the synthesis of mannose-6-phosphate from glucose are normal in carbohydrate deficient glycoprotein syndrome fibroblasts. *Biochemical and biophysical research communications* 208, 517-522.

Pareek, C.S., Smoczynski, R., and Tretyn, A. (2011). Sequencing technologies and genome sequencing. *J Appl Genet* 52, 413-435.

- Park, C., and Zhang, J. (2011). Genome-Wide Evolutionary Conservation of N-Glycosylation Sites. *Molecular Biology and Evolution* 28, 2351-2357.
- Park, J.H., Hogrebe, M., Gruneberg, M., DuChesne, I., von der Heiden, A.L., Reunert, J., Schlingmann, K.P., Boycott, K.M., Beaulieu, C.L., Mhanni, A.A., *et al.* (2015). SLC39A8 Deficiency: A Disorder of Manganese Transport and Glycosylation. *Am J Hum Genet* 97, 894-903.
- Park, M.-H., Rhee, H., Park, J., Woo, H.-M., Choi, B.-O., Kim, B.-Y., Chung, K., Cho, Y.-B., Kim, H., Jung, J.-W., *et al.* (2014). Comprehensive Analysis to Improve the Validation Rate for Single Nucleotide Variants Detected by Next-Generation Sequencing. *PLoS ONE* 9, e86664.
- Perez-Duenas, B., Ormazabal, A., Toma, C., Torrico, B., Cormand, B., Serrano, M., Sierra, C., De Grandis, E., Marfa, M.P., Garcia-Cazorla, A., *et al.* (2011). Cerebral folate deficiency syndromes in childhood: clinical, analytical, and etiologic aspects. *Arch Neurol* 68, 615-621.
- Pérez, B., Medrano, C., Ecay, M., Ruiz-Sala, P., Martínez-Pardo, M., Ugarte, M., and Pérez-Cerdá, C. (2012). A novel congenital disorder of glycosylation type without central nervous system involvement caused by mutations in the phosphoglucomutase 1 gene. *Journal of Inherited Metabolic Disease* 36, 535-542.
- Praeter, C.M., Gerwig, G.J., Bause, E., Nuytinck, L.K., Vliegthart, J.F.G., Breuer, W., Kamerling, J.P., Espeel, M.F., Martin, J.-J.R., Paepe, A.M., *et al.* (2000). A Novel Disorder Caused by Defective Biosynthesis of N-Linked Oligosaccharides Due to Glucosidase I Deficiency. *The American Journal of Human Genetics* 66, 1744-1756.
- Rabbani, B., Tekin, M., and Mahdieh, N. (2014). The promise of whole-exome sequencing in medical genetics. *J Hum Genet* 59, 5-15.
- Rafiq, M., Kuss, A.W., Puettmann, L., Noor, A., Ramiah, A., Ali, G., Hu, H., Kerio, N., Xiang, Y., Garshasbi, M., *et al.* (2011). Mutations in the Alpha 1,2-Mannosidase Gene, MAN1B1, Cause Autosomal-Recessive Intellectual Disability. *The American Journal of Human Genetics* 89, 176-182.
- Ramachandran, N., Munteanu, I., Wang, P., Ruggieri, A., Rilstone, J.J., Israelian, N., Naranian, T., Paroutis, P., Guo, R., Ren, Z.P., *et al.* (2013). VMA21 deficiency prevents vacuolar ATPase assembly and causes autophagic vacuolar myopathy. *Acta Neuropathol* 125, 439-457.
- Raval, K.K., Tao, R., White, B.E., De Lange, W.J., Koonce, C.H., Yu, J., Kishnani, P.S., Thomson, J.A., Mosher, D.F., Ralphe, J.C., *et al.* (2015). Pompe disease results in a Golgi-based glycosylation deficit in human induced pluripotent stem cell-derived cardiomyocytes. *J Biol Chem* 290, 3121-3136.
- Reamon-Buettner, S.M., and Borlak, J. (2010). NKX2-5: an update on this hypermutable homeodomain protein and its role in human congenital heart disease (CHD). *Hum Mutat* 31, 1185-1194.
- Rehm, H.L., Bale, S.J., Bayrak-Toydemir, P., Berg, J.S., Brown, K.K., Deignan, J.L., Friez, M.J., Funke, B.H., Hegde, M.R., Lyon, E., *et al.* (2013). ACMG clinical laboratory standards for next-generation sequencing. *Genetics in Medicine* 15, 733-747.

- Reuter, J.A., Spacek, D.V., Pai, R.K., and Snyder, M.P. (2016). Simul-seq: combined DNA and RNA sequencing for whole-genome and transcriptome profiling. *Nat Methods* **13**, 953-958.
- Reynders, E., Foulquier, F., Teles, E., Quelhas, D., Morelle, W., Rabouille, C., Annaert, W., and Matthijs, G. (2009). Golgi function and dysfunction in the first COG4-deficient CDG type II patient. *Human Molecular Genetics* **18**, 3244-3256.
- Richard, E., Vega, A.I., Pérez, B., Roche, C., Velázquez, R., Ugarte, M., and Pérez-Cerdá, C. (2009). Congenital disorder of glycosylation Ia: New differentially expressed proteins identified by 2-DE. *Biochemical and Biophysical Research Communications* **379**, 267-271.
- Rivinoja, A., Pujol, F.M., Hassinen, A., and Kellokumpu, S. (2012). Golgi pH, its regulation and roles in human disease. *Ann Med* **44**, 542-554.
- Robyns, T., Kuiperi, C., Willems, R., Corveleyn, A., and Nuyens, D. (2015). Targeted capture sequencing in a large LQTS family reveals a new pathogenic mutation c.2038delG in KCNH2 initially missed due to allelic dropout. *Acta Cardiol* **70**, 747-749.
- Rosnoblet, C., Peanne, R., Legrand, D., and Foulquier, F. (2012). Glycosylation disorders of membrane trafficking. *Glycoconjugate Journal* **30**, 23-31.
- S**alvatore, F., Izzo, P., and Paoletta, G. (1986). Aldolase gene and protein families: structure, expression and pathophysiology. *Horiz Biochem Biophys* **8**, 611-665.
- Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463-5467.
- Santer, R., Rischewski, J., von Weihe, M., Niederhaus, M., Schneppenheim, S., Baerlocher, K., Kohlschütter, A., Muntau, A., Posselt, H.G., Steinmann, B., *et al.* (2005). The spectrum of aldolase B (ALDOB) mutations and the prevalence of hereditary fructose intolerance in Central Europe. *Hum Mutat* **25**, 594.
- Sarkozy, A., Hicks, D., Hudson, J., Laval, S.H., Barresi, R., Hilton-Jones, D., Deschauer, M., Harris, E., Rufibach, L., Hwang, E., *et al.* (2013). ANO5 gene analysis in a large cohort of patients with anoctaminopathy: confirmation of male prevalence and high occurrence of the common exon 5 gene mutation. *Hum Mutat* **34**, 1111-1118.
- Scott, K., Gadomski, T., Kozicz, T., and Morava, E. (2014). Congenital disorders of glycosylation: new defects and still counting. *Journal of Inherited Metabolic Disease* **37**, 609-617.
- Schachter, H., and Freeze, H.H. (2009). Glycosylation diseases: quo vadis? *Biochimica et biophysica acta* **1792**, 925-930.
- Schulze, T.G., and McMahon, F.J. (2004). Defining the phenotype in human genetic studies: forward genetics and reverse phenotyping. *Hum Hered* **58**, 131-138.
- Schwarz, J.M., Cooper, D.N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* **11**, 361-362.

- Senderek, J., Müller, J.S., Dusl, M., Strom, T.M., Guergueltcheva, V., Diepolder, I., Laval, S.H., Maxwell, S., Cossins, J., Krause, S., *et al.* (2011). Hexosamine Biosynthetic Pathway Mutations Cause Neuromuscular Transmission Defect. *The American Journal of Human Genetics* *88*, 162-172.
- Shen, J.J., and Chen, Y.T. (2002). Molecular characterization of glycogen storage disease type III. *Curr Mol Med* *2*, 167-175.
- Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B., and Hood, L.E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* *321*, 674-679.
- Smith, R.D., and Lupashin, V.V. (2008). Role of the conserved oligomeric Golgi (COG) complex in protein glycosylation. *Carbohydrate research* *343*, 2024-2031.
- Steinfeld, R., Grapp, M., Kraetzner, R., Dreha-Kulaczewski, S., Helms, G., Dechent, P., Wevers, R., Grosso, S., and Gartner, J. (2009). Folate receptor alpha defect causes cerebral folate transport deficiency: a treatable neurodegenerative disorder associated with disturbed myelin metabolism. *Am J Hum Genet* *85*, 354-363.
- Steinmann, B., and Gitzelmann, R. (1981). The diagnosis of hereditary fructose intolerance. *Helv Paediatr Acta* *36*, 297-316.
- Stray-Pedersen, A., Backe, P.H., Sorte, H.S., Mørkrid, L., Chokshi, N.Y., Erichsen, H.C., Gambin, T., Elgstøen, K.B., Bjørås, M., Wlodarski, M.W., *et al.* (2014). PGM3 mutations cause a congenital disorder of glycosylation with severe immunodeficiency and skeletal dysplasia. *American journal of human genetics* *95*, 96-107.
- Tarailo-Graovac, M., Shyr, C., Ross, C.J., Horvath, G.A., Salvarinova, R., Ye, X.C., Zhang, L.H., Bhavsar, A.P., Lee, J.J., Drogemoller, B.I., *et al.* (2016). Exome Sequencing and the Management of Neurometabolic Disorders. *N Engl J Med* *374*, 2246-2255.
- Tchernitchko, D., Goossens, M., and Wajcman, H. (2004). In silico prediction of the deleterious effect of a mutation: proceed with caution in clinical genetics. *Clinical chemistry* *50*, 1974-1978.
- Tegtmeyer, L.C., Rust, S., van Scherpenzeel, M., Ng, B.G., Losfeld, M.-E.E., Timal, S., Raymond, K., He, P., Ichikawa, M., Veltman, J., *et al.* (2014). Multiple phenotypes in phosphoglucomutase 1 deficiency. *The New England journal of medicine* *370*, 533-542.
- Thiel, C., Meßner-Schmitt, D., Hoffmann, G.F., and Körner, C. (2012). Screening for congenital disorders of glycosylation in the first weeks of life. *Journal of Inherited Metabolic Disease* *36*, 887-892.
- Timal, S., Hoischen, A., Lehle, L., Adamowicz, M., Huijben, K., Sykut-Cegielska, J., Paprocka, J., Jamroz, E., van Spronsen, F.J., Körner, C., *et al.* (2012). Gene identification in the congenital disorders of glycosylation type I by whole-exome sequencing. *Human Molecular Genetics* *21*, 4151-4161.
- Timson, D.J. (2016). The molecular basis of galactosemia - Past, present and future. *Gene* *589*, 133-141.
- Trujillano, D., Perez, B., Gonzalez, J., Tornador, C., Navarrete, R., Escaramis, G., Ossowski, S., Armengol, L., Cornejo, V., Desviat, L.R., *et al.* (2014). Accurate molecular diagnosis of phenylketonuria and tetrahydrobiopterin-deficient hyperphenylalaninemia using high-



throughput targeted sequencing. *Eur J Hum Genet* 22, 528-534.

**U**ngar, D., Oka, T., Krieger, M., and Hughson, F.M. (2006). Retrograde transport on the COG railway. *Trends in cell biology* 16, 113-120.

**V**alayannopoulos, V., Malinova, V., Honzik, T., Balwani, M., Breen, C., Deegan, P.B., Enns, G.M., Jones, S.A., Kane, J.P., Stock, E.O., *et al.* (2014). Sebelipase alfa over 52 weeks reduces serum transaminases, liver volume and improves serum lipids in patients with lysosomal acid lipase deficiency. *J Hepatol* 61, 1135-1142.

Van Den Berghe, G., Hue, L., and Hers, H.G. (1973). Effect of administration of the fructose on the glycogenolytic action of glucagon. An investigation of the pathogeny of hereditary fructose intolerance. *Biochem J* 134, 637-645.

van El, C.G., Cornel, M.C., Borry, P., Hastings, R.J., Fellmann, F., Hodgson, S.V., Howard, H.C., Cambon-Thomsen, A., Knoppers, B.M., Meijers-Heijboer, H., *et al.* (2013). Whole-genome sequencing in health care: recommendations of the European Society of Human Genetics. *Eur J Hum Genet* 21, 580-584.

Van Schaftingen, E., and Jaeken, J. (1995). Phosphomannomutase deficiency is a cause of carbohydrate-deficient glycoprotein syndrome type I. *FEBS letters* 377, 318-320.

Van Scherpenzeel, M., Willems, E., and Lefeber, D.J. (2016). Clinical diagnostics and therapy monitoring in the congenital disorders of glycosylation. *Glycoconj J* 33, 345-358.

Vega, A.I., Pérez-Cerdá, C., Desviat, L.R., Matthijs, G., Ugarte, M., and Pérez, B. (2009). Functional analysis of three splicing mutations identified in the PMM2 gene: toward a new therapy for congenital disorder of glycosylation type Ia. *Human mutation* 30, 795-803.

Vockley, J., Rinaldo, P., Bennett, M.J., Matern, D., and Vladuius, J.D. (2000). Synergistic Heterozygosity: Disease Resulting from Multiple Partial Defects in One or More Metabolic Pathways. *Mol Genet Metab* 71, 10-18.

**W**ada, Y. (2016). Mass spectrometry of transferrin glycoforms to detect congenital disorders of glycosylation: Site-specific profiles and pitfalls. *Proteomics* 16, 3105-3110.

Wierzb-Bobrowicz, T., Lewandowska, E., Stepien, T., and Modzelewska, J. (2008). Immunohistochemical and ultrastructural changes in the brain in probable adult glycogenosis type IV: adult polyglucosan body disease. *Folia Neuropathol* 46, 165-175.

Wolfe, L.A., and Krasnewich, D. (2013). Congenital disorders of glycosylation and intellectual disability. *Developmental Disabilities Research Reviews* 17, 211-225.

Wopereis, S., Grünewald, S., Morava, É., Penzien, J.M., Briones, P., García-Silva, T.M., Demacker, P., Huijben, K., and Wevers, R.A. (2003). Apolipoprotein C-III Isofocusing in the Diagnosis of Genetic Defects in O-Glycan Biosynthesis. *Clinical Chemistry* 49, 1839-1845.

- Wopereis, S., Lefeber, D.J., Morava, E., and Wevers, R.A. (2006). Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review. *Clinical chemistry* 52, 574-600.
- Wu, X., Rush, J.S., Karaoglu, D., Krasnewich, D., Lubinsky, M.S., Waechter, C.J., Gilmore, R., and Freeze, H.H. (2003). Deficiency of UDP-GlcNAc: Dolichol Phosphate N-Acetylglucosamine-1 Phosphate Transferase (DPAGT1) Causes a Novel Congenital Disorder of Glycosylation Type Ij. *Human Mutation* 22, 144-150.
- Wu, X., Steet, R.A., Bohorov, O., Bakker, J., Newell, J., Krieger, M., Spaapen, L., Kornfeld, S., and Freeze, H.H. (2004). Mutation of the COG complex subunit gene COG7 causes a lethal congenital disorder. *Nature Medicine* 10, 518-523.
- Würde, A.E., Reunert, J., Rust, S., Hertzberg, C., Haverkämper, S., Nürnberg, G., Nürnberg, P., Lehle, L., Rossi, R., and Marquardt, T. (2012). Congenital disorder of glycosylation type Ij (CDG-Ij, DPAGT1-CDG): Extending the clinical and molecular spectrum of a rare disease. *Molecular Genetics and Metabolism* 105, 634-641.
- X**ue, Y., Ankala, A., Wilcox, W.R., and Hegde, M.R. (2015). Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. *Genet Med* 17, 444-451.
- Y**ang, A.C., Ng, B.G., Moore, S.A., Rush, J., Waechter, C.J., Raymond, K.M., Willer, T., Campbell, K.P., Freeze, H.H., and Mehta, L. (2013). Congenital disorder of glycosylation due to DPM1 mutations presenting with dystroglycanopathy-type congenital muscular dystrophy. *Molecular Genetics and Metabolism* 170, 345-351.
- Yuste-Checa, P., Brasil, S., Gamez, A., Underhaug, J., Desviat, L.R., Ugarte, M., Perez-Cerda, C., Martinez, A., and Perez, B. (2016). Pharmacological Chaperoning: A Potential Treatment for PMM2-CDG. *Hum Mutat*, In press.
- Z**eevaert, R., Foulquier, F., Cheillan, D., Cloix, I., Guffon, N., Sturiale, L., Garozzo, D., Matthijs, G., and Jaeken, J. (2009). A new mutation in COG7 extends the spectrum of COG subunit deficiencies. *European journal of medical genetics* 52, 303-305.
- Zhang, W., He, Y., Wang, W., Han, Z., He, J., Chen, N., Dong, L., Tao, S., He, J., Zhu, F., *et al.* (2015). HLA-A locus allelic dropout in Sanger sequence-based typing due to the single nucleotide polymorphism of exon 1. *Int J Immunogenet* 42, 457-460.
- Zühlsdorf, A., Park, J., Wada, Y., Rust, S., Reunert, J., DuChesne, I., Grüneberg, M., and Marquardt, T. (2015). Transferrin variants: Pitfalls in the diagnostics of Congenital disorders of glycosylation. *Clinical Biochemistry* 48, 11-13.
- Zutt, R., van der Kooi, A.J., Linthorst, G.E., Wanders, R.J., and de Visser, M. (2014). Rhabdomyolysis: review of the literature. *Neuromuscul Disord* 24, 651-659.



ANEXOS

**Anexo 1. Relación de genes implicados en los defectos en la N-glicosilación y en los defectos combinados.**

Gen	MIM	Función	Enfermedad	Herencia
<b>Defectos combinados: Síntesis de azúcares unidos a nucleótidos</b>				
<i>GALE</i>	606953	UDP-galactosa 4-epimerasa	Galactosemia	AR
<i>GALK1</i>	604313	Galactosa kinasa	Galactosemia	AR
<i>GALT</i>	606999	Galatosa-1-P uridiltransferasa	Galactosemia	AR
<i>GFPT1</i>	138292	Glutamina:fructosa-6-fosfato amidotransferasa	GFPT1-CDG, síndrome congénito miasténico	AR
<i>GMPPA</i>	615495	Subunidad alfa de GDP-manosa pirofosforilasa	GMPPA-CDG	AR
<i>GMPPB</i>	615320	Subunidad beta de GDP-manosa fosforilasa	GMPPB-CDG	AR
<i>GNE</i>	603824	UDP-N-acetilglucosamina 2-epimerasa	Miopatía hereditaria con cuerpos de inclusión	AR / AD
<i>NANS</i>	605202	Ácido N-acetilneuramínico-6-P sintasa	Displasia	AR
<i>MPI</i>	154550	Fosfomanosa isomerasa	MPI-CDG	AR
<i>PGM1</i>	171900	Fosfoglucomutasa 1	PGM1-CDG, GSD 14	AR
<i>PGM3</i>	172100	Fosfoglucomutasa 3	Inmunodeficiencia 23	AR
<i>PMM2</i>	601785	Fosfomanomutasa	PMM2-CDG	AR
<b>Defectos combinados: Homeostasis del AG</b>				
<i>ATP6AP1</i>	300197	Factor de ensamblaje de la ATPasa vacuolar / tráfico vesicular	Inmunodeficiencia	XLR
<i>ATP6VOA2</i>	611716	Bomba de protones ATPasa del AG	ATP6VOA2-CDG, cutis laxa autosómica recesiva tipo 2	AR
<i>COG1</i>	606973	Subunidad 1 de COG	COG1-CDG	AR
<i>COG2</i>	606974	Subunidad 2 de COG	COG2-CDG	AR
<i>COG4</i>	606976	Subunidad 4 de COG	COG4-CDG	AR
<i>COG5</i>	606821	Subunidad 5 de COG	COG5-CDG	AR
<i>COG6</i>	606977	Subunidad 6 de COG	COG6-CDG	AR
<i>COG7</i>	606978	Subunidad 7 de COG	COG7-CDG	AR
<i>COG8</i>	606979	Subunidad 8 de COG	COG8-CDG	AR
<i>SEC23B</i>	610512	Subunidad de COP-II	Anemia deseritropoyética congénita tipo 2	AR / AD
<i>SLC35A1</i>	605634	Transportador de CMP-ácido siálico al AG	SLC35A1-CDG	AR
<i>SLC35A2</i>	314375	Transportador de UDP-galactosa al AG	SLC35A2-CDG	XLD
<i>SLC35C1</i>	605881	Transportador de GDP-fucosa al AG	SLC35C1-CDG	AR
<i>SLC35D1</i>	610804	Transportador de UDP-GlcA/NAcGal al AG	Displasia de Schneckenbecken	AR
<i>TMEM165</i>	614726	Antiportador $\text{Ca}^{+2}/\text{H}^{+}$ del AG	TMEM165-CDG	AR
<i>TMEM199</i>	616815	Factor de ensamblaje de la ATPasa vacuolar / tráfico vesicular	TMEM199-CDG	AR
<i>VMA21</i>	300913	Factor de ensamblaje de la ATPasa vacuolar / tráfico vesicular	Miopatía ligada al X	XLR

**Anexo 1 (continuación 1). Relación de genes implicados en los defectos en la N-glicosilación y en los defectos combinados.**

Gen	MIM	Función	Enfermedad	Herencia
<b>Defectos combinados: Síntesis de dolicol unido a azúcares</b>				
<i>DHDDS</i>	608172	Cis-prenol sintasa	Retinitis pigmentosa	AR
<i>DOLK</i>	610746	Dolicol kinasa	DOLK-CDG	AR
<i>DPM1</i>	603503	Subunidad 1 de Dol-P Man sintasa	DPM1-CDG	AR
<i>DPM2</i>	603564	Subunidad 2 de Dol-P Man sintasa	DPM2-CDG	AR
<i>DPM3</i>	605951	Subunidad 3 de Dol-P Man sintasa	DPM3-CDG	AR
<i>MPDU1</i>	604041	Dol-P + GDP-Man/UDP-Glc → Dol-P-Man/Glc	MPDU1-CDG	AR
<i>SRD5A3</i>	611715	Poliprenol reductasa	SRD5A3-CDG	AR
<b>N-glicosilación: Biosíntesis de oligosacárido unido a dolicol</b>				
<i>ALG1</i>	605907	β1-4 Manosiltransferasa	ALG1-CDG	AR
<i>ALG2</i>	607905	α1-3/6 Manosiltransferasa	ALG2-CDG	AR
<i>ALG3</i>	608750	α1-3 Manosiltransferasa	ALG3-CDG	AR
<i>ALG6</i>	604566	α1-3 Glucosiltransferasa	ALG6-CDG	AR
<i>ALG8</i>	608103	α1-3 Glucosiltransferasa	ALG8-CDG	AR
<i>ALG9</i>	606941	α1-2 Manosiltransferasa	ALG9-CDG	AR
<i>ALG11</i>	613666	α1-2 Manosiltransferasa	ALG11-CDG	AR
<i>ALG12</i>	607144	α1-6 Manosiltransferasa	ALG12-CDG	AR
<i>ALG13</i>	300776	UDP-NAcGlc transferasa	ALG13-CDG	AR
<i>ALG14</i>	612866	UDP-NAcGlc transferasa	ALG14-CDG	AR
<i>DPAGT1</i>	191350	NAcGlc-1-P transferasa	DPAGT1-CDG	AR
<i>RFT1</i>	611908	Flipasa	RFT1-CDG	AR
<b>N-glicosilación: Complejo oligosacariltransferasa (OST)</b>				
<i>DDOST</i>	602202	Oligosacariltransferasa	DDOST-CDG	AR
<i>MAGT1</i>	300715	Oligosacariltransferasa	Retraso mental ligado al X	XLR
<i>SSR4</i>	300090	Subunidad 4 del complejo TRAP	SSR4-CDG	XLR
<i>STT3A</i>	601134	Oligosacariltransferasa	STT3A-CDG	AR
<i>STT3B</i>	608605	Oligosacariltransferasa	STT3B-CDG	AR
<i>TUSC3</i>	601385	Oligosacariltransferasa	Retraso mental AR	AR
<b>N-glicosilación: Procesamiento del oligosacárido unido a proteína</b>				
<i>B4GALT1</i>	137060	β1-4 Gal-transferasa	B4GALT1-CDG	AR
<i>GANAB</i>	104160	Subunidad 2 de glucosidasa II	Riñón poliquístico tipo 3	AD
<i>MAN1B1</i>	604346	α1-2 manosidasa I	Retraso mental autosómico recesivo	AR
<i>MGAT2</i>	602616	β1-2 NAcGlc-transferasa	MGAT2-CDG	AR
<i>MOGS</i>	601336	α1-2 glucosidasa	MOGS-CDG	AR
<i>ST3GAL3</i>	606494	α2-3 Sia-transferasa	Retraso mental AR	AR

AD, autosómica dominante; AR, autosómica recesiva; XLD, ligada al X dominante; XLR, ligada al X recesiva.

## Anexo 2. Relación de genes implicados en la glucogenosis.

Gen	MIM	Función	Enfermedad	Herencia
<b>Glucogenosis hepática</b>				
<i>G6PC</i>	613742	Glucosa-6-fosfatasa	GSD 1A, enfermedad de Von Gierke	AR
<i>GYS2</i>	138571	Glucógenos sintasa 2	GSD 0a	AR
<i>PHKA2</i>	300798	Subunidad $\alpha$ de fosforilasa kinasa hepática	GSD 9a	XLR
<i>PHKG2</i>	172471	Subunidad $\gamma$ de fosforilasa kinasa hepática	GSD 9c	AR
<i>PYGL</i>	613741	Glucógeno fosforilasa hepática	GSD 6, enfermedad de Hers	AR
<i>SLC37A4</i>	602671	Glucosa-6-fosfato translocasa	GSD 1b	AR
<b>Glucogenosis muscular</b>				
<i>ENO3</i>	131370	Enolasa 3 muscular	GSD 13	AR
<i>GYG1</i>	603942	Glucogenina 1	GSD 15	AR
<i>GYS1</i>	138570	Glucógeno sintasa 1	GSD 0b	AR
<i>LDHA</i>	150000	Lactato deshidrogenasa A	GSD 11	AR
<i>PFKM</i>	610681	Fosfofructokinasa muscular	GSD 7, enfermedad de Tauri	AR
<i>PGAM2</i>	612931	Fosfoglicerato mutasa 2 muscular	GSD 10	AR
<i>PGM1</i>	171900	Fosfoglucomutasa 1	GSD 14, PGM1-CDG	AR
<i>PHKA1</i>	311870	Subunidad $\alpha$ de fosforilasa kinasa muscular	GSD 9d	XLR
<i>PYGM</i>	618455	Glucógeno fosforilasa muscular	GSD 5, enfermedad de Mc Ardel	AR
<b>Glucogenosis generalizada</b>				
<i>AGL</i>	610860	Amilo $\alpha$ -1,6-glucosidasa (enzima desramificante)	GSD 3, enfermedad de Cori-Forbes	AR
<i>GAA</i>	606800	$\alpha$ -1,4-glucosidasa ácida lisosomal	GSD 2a, enfermedad de Pompe	AR
<i>GBE1</i>	607839	Enzima ramificante del glucógeno	GSD 4, enfermedad de Andersen	AR
<i>LAMP2</i>	309060	Glicoproteína de la membrana lisosomal	GSD2b, enfermedad de Danon	XLD
<i>PHKB</i>	172490	Subunidad $\beta$ de fosforilasa kinasa	GSD 9b	AR
<b>Otras glucogenosis</b>				
<i>ALDOA</i>	103850	Fructosa-1,6-bifosfato aldolasa A eritrocitaria y muscular	GSD 12	AR
<i>EPM2A</i>	607566	Laforina	Enfermedad de Lafora	AR
<i>NHLRC1</i>	608072	Malina	Enfermedad de Lafora	AR
<i>G6PC3</i>	611045	Glucosa-6-fosfatasa 3	Neutropenia congénita severa, síndrome de Dursun	AR
<i>PGK1</i>	311800	Fosfoglicerato kinasa 1	Deficiencia en <i>PGK1</i>	XLR
<i>PRKAG2</i>	602743	Proteína kinasa activada por AMP	GSD H, Cardiomiopatía hipertrófica familiar 6 (CMH6)	AD
<i>SLC2A2</i>	138160	Transportador de glucosa	Síndrome de Fanconi-Bickel	AR

AD, autosómica dominante; AR, autosómica recesiva; XLD, ligada al X dominante; XLR, ligada al X recesiva.

### Anexo 3. Genes capturados en el panel de CDG (TES-CDG).

Gen	Número de transcrito	Locus
<b>Defectos en múltiples rutas de glicosilación: Síntesis de azúcares unidos a nucleótidos</b>		
<i>GNE</i>	NM_005476	9p13.3
<i>MPI</i>	NM_002435	15q24.1
<i>PMM2</i>	NM_000303	16p13.2
<b>Defectos en múltiples rutas de glicosilación: Homeostasis del AG</b>		
<i>ATP6VOA2</i>	NM_012463	12q24.31
<i>COG1</i>	NM_018714	17q25.1
<i>COG2</i>	NM_007357	1q42.2
<i>COG3</i>	NM_031431	13q14.13
<i>COG4</i>	NM_015386	16q22.1
<i>COG5</i>	NM_006348	7q22.3
<i>COG6</i>	NM_020751	13q14.11
<i>COG7</i>	NM_153603	16p12.2
<i>COG8</i>	NM_032382	16q22.1
<i>SEC23B</i>	NM_006363	20p11.23
<i>SLC35A1</i>	NM_006416	6q15
<i>SLC35C1</i>	NM_018389	11p11.2
<b>Defectos en múltiples rutas de glicosilación: Síntesis de dolicol unido a azúcares</b>		
<i>DHDDS</i>	NM_024887	1p36.11
<i>DOLK</i>	NM_014908	9q34.11
<i>DPM1</i>	NM_003859	20q13.13
<i>DPM2</i>	NM_003863	9q34.11
<i>DPM3</i>	NM_153741	1q22
<i>MPDU1</i>	NM_004870	17p13.1
<i>SRD5A3</i>	NM_024592	4q12
<b>N-glicosilación: Biosíntesis de oligosacárido unido a dolicol</b>		
<i>ALG1</i>	NM_019109	16p13.3
<i>ALG2</i>	NM_033087	9q22.33
<i>ALG3</i>	NM_005787	3q27.1
<i>ALG6</i>	NM_013339	1p31.3
<i>ALG8</i>	NM_024079	11q14.1
<i>ALG9</i>	NM_024740	11q23.1
<i>ALG11</i>	NM_001004127	13q14.3
<i>ALG12</i>	NM_024105	22q13.33
<i>DPAGT1</i>	NM_001382	11q23.3
<i>RFT1</i>	NM_052859	3p21.1
<b>N-glicosilación: Complejo oligosacariltransferasa (OST)</b>		
<i>DAD1</i>	NM_001344	14q11.2
<i>DDOST</i>	NM_005216	1p36.12
<i>MAGT1</i>	NM_032121	Xq21.1
<i>RPN1</i>	NM_002950	3q21.3
<i>RPN2</i>	NM_002951	20q11.23
<i>STT3A</i>	NM_152713	11q24.2
<i>STT3B</i>	NM_178862	3p23
<i>TUSC3</i>	NM_006765	8p22
<b>N-glicosilación: Procesamiento del oligosacárido unido a proteína</b>		
<i>B4GALT1</i>	NM_001497	9p21.1
<i>MGAT2</i>	NM_002408	14q21.3
<i>MOGS</i>	NM_006302	2p13.1

**Anexo 4. Genes de glucogenosis capturados en el panel metabólico (TES-EMH) y priorizados en el análisis de variantes de CES.**

Gen	Número de transcrito	Locus
<b>Glucogenosis hepática</b>		
<i>G6PC</i>	NM_000151	17q21.31
<i>GYS2</i>	NM_021957	12p12.1
<i>PHKA2</i>	NM_000292	Xp22.13
<i>PHKG2</i>	NM_000294	16p11.2
<i>PYGL</i>	NM_002863	14q22.1
<i>SLC37A4</i>	NM_001164277	11q23.3
<b>Glucogenosis muscular</b>		
<i>ENO3</i>	NM_053013	17p13.2
<i>GYG1</i>	NM_004130	3q24
<i>GYS1</i>	NM_002103	19q13.33
<i>LDHA</i>	NM_005566	11p15.1
<i>PFKM</i>	NM_000289	12q13.11
<i>PGAM2</i>	NM_000290	7p13
<i>PGM1</i>	NM_002633	1p31.3
<i>PHKA1</i>	NM_002637	Xq13.1
<i>PHKG1</i>	NM_001258459	7p11.2
<i>PYGM</i>	NM_005609	11q13.1
<b>Glucogenosis generalizada</b>		
<i>AGL</i>	NM_000642	1p21.2
<i>GAA</i>	NM_000152	17q25.3
<i>GBE1</i>	NM_000158	3p12.2
<i>PHKB</i>	NM_000293	16q12.1
<b>Otras glucogenosis</b>		
<i>ALDOA</i>	NM_000034	16p11.2
<i>SLC2A2</i>	NM_000340	3q26.2

## **Anexo 5. Listado priorizado de genes CDG utilizado para el análisis por CES y WES.**

Biosíntesis y reciclado de dolicol-fosfato: *ACAT2* (NM\_005891), *DHDDS* (NM\_024887), *DOLK* (NM\_014908), *DOLPP1* (NM\_020438), *FDPS* (NM\_001135821), *GGPS1* (NM\_001037277), *HMGCR* (NM\_000859), *HMGCS1* (NM\_002130), *IDI1* (NM\_004508), *IDI2* (NM\_033261), *MVD* (NM\_002461), *MVK* (NM\_000431), *NUS1* (NM\_138459), *PMVK* (NM\_006556), *SRD5A3* (NM\_024592).

Síntesis de azúcares activados: *CMAS* (NM\_018686), *FPTG* (NM\_003838), *FUK* (NM\_145059), *GALE* (NM\_00403), *GALK1* (NM\_000154), *GALK2* (NM\_001001556), *GALM* (NM\_138801), *GALT* (NM\_000155), *GCK* (NM\_000162), *GFPT1* (NM\_002056), *GMDS* (NM\_001500), *GMPPA* (NM\_205847), *GMPPB* (NM\_021971), *GNE* (NM\_005476), *GNPDA1* (NM\_005471), *GNPNAT1* (NM\_198066), *GPI* (NM\_000175), *HK3* (NM\_002115), *MPI* (NM\_002435), *NANP* (NM\_152667), *NANS* (NM\_018946), *NPL* (NM\_030769), *PGM1* (NM\_002633), *PGM3* (NM\_015599), *PMM2* (NM\_000303), *RENBP* (NM\_002910), *TSTA3* (NM\_003313), *UAP1* (NM\_003115), *UGDH* (NM\_003359), *UGP2* (NM\_006759), *UXS1* (NM\_025076).

Síntesis del oligosacárido unido a dolicol y transferencia a la proteína: *ALG1* (NM\_019109), *ALG2* (NM\_033087), *ALG5* (NM\_013338), *ALG11* (NM\_001004127), *ALG13* (NM\_001099922), *ALG14* (NM\_144988), *DPAGT1* (NM\_001382), *DPM1* (NM\_003859), *DPM2* (NM\_003863), *DPM3* (NM\_153741), *MPDU1* (NM\_004870), *RFT1* (NM\_052859), *ALG3* (NM\_005787), *ALG6* (NM\_013339), *ALG8* (NM\_024079), *ALG9* (NM\_024740), *ALG10* (NM\_032834), *ALG10B* (NM\_001013620), *ALG12* (NM\_024105), *DAD1* (NM\_001344), *DDOST* (NM\_005216), *MAGT1* (NM\_032121), *OST4* (NM\_001134693), *RPN1* (NM\_002950), *RPN2* (NM\_002951), *SEC61A1* (NM\_001142628), *SEC61A2* (NM\_018144), *SEC61B* (NM\_006808), *SEC61G* (NM\_014302), *SSR1* (NM\_003144), *SSR2* (NM\_003145), *SSR3* (NM\_007107), *SSR4* (NM\_006280), *STT3A* (NM\_152713), *STT3B* (NM\_178862), *TUSC3* (NM\_006765).

Modificación del oligosacárido unido a la proteína: *B1GALT1* (NM\_001497), *B4GALT2* (NM\_003780), *B4GALT3* (NM\_001199874), *GANAB* (NM\_198335), *MAN1A1* (NM\_005907), *MAN1A2* (NM\_006699), *MAN1B1* (NM\_016219), *MAN1C1* (NM\_020379), *MAN2A1* (NM\_002372), *MAN2A2* (NM\_006122), *MAN2B1* (NM\_000528), *MAN2B2* (NM\_015274), *MAN2C1* (NM\_001256494), *MANBA* (NM\_005908), *MGAT1* (NM\_001114618), *MGAT2* (NM\_002408), *MGAT3* (NM\_002409), *MOGS* (NM\_006302), *ST3GAL3* (NM\_006279).

Proteínas de la homeostasis del AG: *ATP6VOA2* (NM\_012463), *COG1* (NM\_018714), *COG2* (NM\_007357), *COG3* (NM\_031431), *COG4* (NM\_015386), *COG5* (NM\_006348), *COG6* (NM\_020751), *COG7* (NM\_153603), *COG8* (NM\_032382), *SEC23B* (NM\_006363), *SLC35A1* (NM\_006416), *SLC35A2* (NM\_005660), *SLC35C1* (NM\_018389), *SLC35D1* (NM\_015139), *TMEM165* (NM\_018475).

Otras proteínas: *AGA* (NM\_000027), *B4GALNT3* (NM\_173593), *B4GALNT4* (NM\_178537), *CHST8* (NM\_022467), *CHST9* (NM\_031422), *ENGASE* (NM\_001042573), *FUCA1* (NM\_000147), *FUCA2* (NM\_032020), *FUT8* (NM\_178155), *GALNT8* (NM\_017417), *GLB1* (NM\_000404), *HEXA* (NM\_005020), *HEXB* (NM\_000521), *HEXDC* (NM\_173620), *MGAT4A* (NM\_012214), *MGAT4B* (NM\_014275), *MGAT4C* (NM\_013244), *MGAT5* (NM\_002410), *MGAT5B* (NM\_144677), *NEU1* (NM\_000434), *NEU2* (NM\_005383), *NEU3* (NM\_006656), *NEU4* (NM\_080741), *NGLY1* (NM\_018297), *RER1* (NM\_007033), *SLC2A1* (NM\_006516), *SLC5A9* (NM\_001135181), *ST6GAL1* (NM\_173216), *ST6GAL2* (NM\_032528), *ST6GALNAC1* (NM\_018414).



# PUBLICACIONES

1	PUBLICACIONES QUE HAN DADO LUGAR A ESTA TESIS	143
2	OTROS TRABAJOS PUBLICADOS DURANTE LA ETAPA PREDOCTORAL	199

## 1 PUBLICACIONES QUE HAN DADO LUGAR A ESTA TESIS

Pérez-Cerdá, C., Girós, M.L., Serrano, M., Ecay, M.J., Gort, L., Pérez Dueñas, B., **Medrano, C.**, García Alix, A., Artuch, R., Briones, P., Pérez, B.. (2017) A population based study on congenital disorders of protein N-glycosylation experience in clinical and genetic diagnosis. En prensa.

Jansen, J.C., Cirak, S., van Scherpenzeel, M., Timal, S., Reunert, J., Rust, S., Perez, B., Vicogne, D., Krawitz, P., Wada, Y., Ashikov, A., Perez-Cerda, C., **Medrano, C.**, Arnoldy, A., Hoischen, A., Huijben, K., Steenbergen, G., Quelhas, D., Diogo, L., Rymen, D., Jaeken, J., Guffon, N., Cheillan, D., van den Heuvel, L.P., Maeda, Y., Kaiser, O., Schara, U., Gerner, P., van den Boogert, M.A., Holleboom, A.G., Nassogne, M.C., Sokal, E., Salomon, J., van den Bogaart, G., Drenth, J.P., Huynen, M.A., Veltman, J.A., Wevers, R.A., Morava, E., Matthijs, G., Foulquier, F., Marquardt, T., Lefeber, D.J.. (2016). CCDC115 Deficiency Causes a Disorder of Golgi Homeostasis with Abnormal Protein Glycosylation. *Am J Hum Genet.* 98(2), 310-21.

Vega, A.I.\*, **Medrano, C.\***, Navarrete, R., Desviat, L.R., Merinero, B., Rodriguez-Pombo, P., Vitoria, I., Ugarte, M., Perez-Cerda, C., Perez, B., (2016). Molecular diagnosis of glycogen storage disease and disorders with overlapping clinical symptoms by massive parallel sequencing. *Genet Med.* 18(10), 1037-43.

Pérez-Cerdá C., Girós M.L., Serrano M., Pérez Dueñas B., Ecay M.J., **Medrano C.**, Gort L., Pérez B., (2015). *Acta Pediatr Esp.* Protocolo de diagnóstico y tratamiento de los defectos congénitos de glicosilación 73(Supl.), 13-27.

Vega, A.I.\*, **Medrano, C.\***, Navarrete, R., Desviat, L.R., Merinero, B., Rodriguez-Pombo, P., Vitoria, I., Ugarte, M., Perez-Cerda, C., Perez, B., (2016). Diagnóstico molecular por secuenciación masiva de glucogenosis y enfermedades con síntomas clínicos solapantes. *Genética Médica.*

## 2 OTROS TRABAJOS PUBLICADOS DURANTE LA ETAPA PREDOCTORAL \_\_\_\_\_

Yuste-Checa, P., **Medrano, C.**, Gámez, A., Desviat, L.R., Matthijs, G., Ugarte, M., Pérez-Cerdá, C., Pérez, B., (2015). Antisense-mediated therapeutic pseudoexon skipping in TMEM165-CDG. *Clinical Genetics* 87, 42-48.

Pérez, B., **Medrano, C.**, Ecay, M., Ruiz-Sala, P., Martínez-Pardo, M., Ugarte, M., Pérez-Cerdá, C., (2012). A novel congenital disorder of glycosylation type without central nervous system involvement caused by mutations in the phosphoglucomutase 1 gene. *Journal of Inherited Metabolic Disease* 36, 535-542.